RADC-TR-65-124
Final Report

A SURVEY AND EXTENSION OF TECHNIQUES IN FACTOR ANALYSIS

Larry E. Irwin
Elisabeth B. Ostermann
et al

TECHNICAL REPORT NO. RADC-TR-65-124
August 1965

Information Processing Branch
Rome Air Development Center
Research and Technology Division
Air Force Systems Command
Griffiss Air Force Base, New York

A SURVEY AND EXTENSION OF TECHNIQUES IN FACTOR ANALYSIS

Larry E. Irwin
Elisabeth B. Ostermann
et al.

AFLC, GAFB, N.Y., 10 Sep 65-111

FOREWORD

This final report was prepared by Systems Research Laboratories, Inc., Dayton, Ohio, under Contract No. AF30(602)-3300. The contract was initiated under Project No. 5581, and Task No. 558110. The report describes the result of a year of study conducted under the direction of Mr. Rocco F. Iuorno, Rome Air Development Center (EMIIH), Research and Technology Division.

This report covers work conducted from February 1964 to February 1965. The authors include David H. Brand: Sections I; 5.2; 5.3., Dr. Thomas G. Donnelly: Sections 6.2; 6.3; Appendix III., James N. Duley: Section 7.2., and Jeff C. Shoemacher: Section 5.4. The authors of all other sections are Larry E. Irwin and Elisabeth B. Ostermann.

David H. Brand was project manager from February 1964 to January 1965.

Acknowledgement for contributions is given to Janet R. Gum and Robert M. Linhart.

Consultants on the contract were Drs. Raymond B. Cattell and Herbert W. Eber.

This technical report has been reviewed and is approved.

Approved:
FRANK J. TOMAINI
Chief, Information Processing Branch
Intel and Info Processing Division

Approved:
ROBERT J. QUINN, JR.
Colonel, USAF
Chief, Intel and Info Processing Div

FOR THE COMMANDER:
IRVING A. GABELMAN
Chief, Advanced Studies Group

ABSTRACT

The purpose of the work reported here was to present the structure of factor analysis to a physical scientist and to extend the structure where it was weakest.

The reference guide in the appendix performs as a dynamic survey of factor analysis by guiding a neophyte factor analyst through an application. Reference is made to expanded presentations in the body of the report.

The structure of factor analysis has been extended in the following areas: effects of the number of observations, sampling effect, interpretation of factors, and communality.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

LIST OF FIGURES

LIST OF TABLES

EVALUATION

The purpose of this work was to study techniques in factor analysis in order to provide an objective and mathematical standard in the field. This study was needed to make factor analysis a useful analytical tool for practicing engineers and scientists. Those areas investigated which have made factor analysis less attractive for use as an analytic tool are: problem of communality estimates, number of observations for a valid factor analysis, uniqueness, and sampling effects on factor structure. Attempts were made and were partially successful in storing these problems. The results of this study are two fold:

(1) An attempt to explain mathematically the events occurring during a factor analysis which can be understood by engineers and scientists. This in turn will allow a practicing engineer to make an objective decision whether he can use factor analysis as an analytic tool.

(2) Once an engineer decides to use factor analysis in his work, a handbook or reference guide is provided which outlines a step by step procedure for conducting a factor analysis; starting with the construction of his experiment and ending with aids to interpret results. Computer program descriptions are also provided including formats for inputting raw data.

The results of this study have already been put to practice by members of EMIIH in constructing an experimental classification model to be used for automatic dissemination of technical documents to engineers and scientists in RADC.

R. Iuorno

RADC Project Engineer

Section I

INTRODUCTION

1.1 BRIEF HISTORY

It is appropriate to begin the Introduction to this final report with Truman Kelley's remarks made in his 1940 publication (Reference 1, p. 120):

> "There is no search for timeless, spaceless, population-less truth in factor analysis; rather, it represents a simple, straightforward problem of description in several dimensions of a definite group functioning in definite manners, and he who assumes to read more remote verities into the factorial outcome is certainly doomed to disappointment."

This particular passage was also selected by Harman (Reference 2, p. 5) to emphasize the simplicity of the problem and potential pitfalls of understanding its solution. Regardless of what is done in methodology or conceptual studies, an acceptance of the basic model necessarily implies that the problem remains simple and the solution remains ambiguous.

Since factor analysis was found useful around the turn of this century by a psychologist, Charles Spearman, and described mathematically by a statistician, Karl Pearson, the development of techniques has more or less followed the lines of the empirical school. That is, methods to obtain factor solutions have evolved more from the necessity of describing certain underlying psychological entities by meaningful groups of hypothetical constructs than from an application of advanced mathematical ideas to the basic mathematical problem. As a consequence, factor analysis suffered from a lack of mathematical ordering of its esoteric devices until Harry Harman, in close association with Karl Holzinger, published in 1960 an excellent summary of most of the significant factor analysis work which had been done to that time (Reference 2). This book, Modern Factor Analysis, has been welcomed

into and accepted by most of the factor analysis groups in this country as a general reference guide useful in selecting an appropriate method or set of methods. Its comparative presentations are very good.

The intent of this study was not simply a reiteration of Harman's work with, perhaps, a few more up-to-date details. Rather it was an investigation into a few of the unsolved, classic mathematical problems with a demonstration of how too little knowledge of necessary assumptions concerning these problems can be troublesome and at times devastating. Attempts were made and were partially successful in solving the problems of communality estimates, number of observations for a valid factor analysis, uniqueness, and sampling error effects on factor structure.

1.2 THE MODEL AND SOME MATRIC NOTATIONS

Factor analysis is concerned with the study of an array of numbers which has certain properties and contains information about linear relationships among sets of data points. This array is called a correlation matrix and the numbers, or entries, are called correlation coefficients. The array is so constructed that the number in the i^{th} row and j^{th} column represents the correlation, or degree of linear relationship ($y = ax + b$), between the i^{th} and j^{th} sets of data points. For 5 sets of data points such an array might look like:

	1	2	3	4	5
1	-	.6	.4	.3	.1
2	.6	-	.1	-.2	-.6
3	.4	.1	-	.2	.1
4	.3	-.2	.2	-	.8
5	.1	-.6	.1	.8	-

Easily noted is that the number in the 4th row and 5th column is the same as the number in the 5th row and 4th column, and, in fact, the number in the i^{th} row and j^{th} column (call it r_{ij}) is the same

as the number in the j^{th} row and i^{th} column (call it r_{ji}). This property of symmetry, as well as others, will be stated more formally a little later but is worth noting in a preliminary discussion on the classic problems and the model.

The problems treated in this report are mostly those which have caused mathematicians to reject factor analysis as a useful analytical tool. Many of the reasons for rejection are unjustified--some are justified. Those reasons which are unjustified concern the misunderstanding or misuse of the basic model and/or assumptions necessary in determining a "unique" solution.

The basic model stated simply is this: given a correlation matrix for a set of data points with appropriately selected diagonal values, determine a set of factors (or hypothetical variables) which when linearly combined reproduce the original set of data points. In a sense, then, the model is the same as for multiple linear regression only the independent variables are replaced by hypothetical variables. The big difference, of course, is that the final synthesis of original data points is complete for all variables in factor analysis and complete only for the dependent variables in a regression.

Let us adopt the vector notation X to mean an ordered sequence of values, or elements, $(x_1, x_2, \ldots, x_N)$. Then in vector notation the linearity of the model is seen to be

$$X_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + a_jU_j$$

where X_j is the original set of observations, F_1 through F_m are the hypothetical common variables, or factors, U_j is the unique factor, and the coefficients a_{j1} through a_{jm} and a_j are those loadings required to reproduce X_j. The "linearity" of the model can not be overemphasized. In most multivariate studies, it is at best a crude approximation to inherent nonlinearities which occur in nature. The model represents a compromise between synthesis accuracy and computational feasibility, a compromise which is too often considered inviolate for interpretation purposes.

Before outlining the classic problems let us digress an instant to review some matrix algebra and notation. A set of vectors arranged in such a manner that the elements of the vectors form rows and columns is called a matrix and will be denoted by a capital letter, e.g. R. To illustrate further using the previous example:

$$R = \begin{pmatrix} 1.0 & .6 & .4 & .3 & .1 \\ .6 & 1.0 & .1 & -.2 & -.6 \\ .4 & .1 & 1.0 & .2 & .1 \\ .3 & -.2 & .2 & 1.0 & .8 \\ .1 & -.6 & .1 & .8 & 1.0 \end{pmatrix} = (r_{ij})$$

Note that the diagonal elements are ones, a classic problem we shall dwell on shortly. The transpose R^T of this matrix is simply the matrix with its rows and columns interchanged such that a typical element r_{ij} becomes r_{ji}. A symmetric matrix is a matrix which is the same as its transpose-- $R = R^T$ in our example.

To review the four fundamental matrix operations:

$$A + B = (a_{ij} + b_{ij})$$

$$A - B = (a_{ij} - b_{ij})$$

$$AB = \left(\sum_{k=1}^{N} a_{ik} b_{kj} \right) \quad \text{where } N \text{ is the number of columns in } A \text{ and rows in } B$$

$$cA = (ca_{ij})$$

The row order of a matrix is the number of rows of the matrix. The column order is the number of columns. The determinant of a matrix of order N is the summation defined as follows:

$$\det A = |A| = \sum_{k=1}^{N} (-1)^{i+k} a_{ik} \det A(i|k)$$

where $1 \leq i \leq N$ and $A(i|k)$ denotes a matrix with the i^{th} row and k^{th} column removed. Starting with the determinant of a second order square matrix (number of rows equals the number of columns) the idea of using a determinant to define a determinant presents itself as being the easiest to understand and illustrates the difficulty of deriving another definition. A matrix is singular if $\det A = 0$.

Some important theorems in applying matrix theory to factor analysis are:

Theorem 1.1: $\det A = \det A^T$

Theorem 1.2: If all elements of any column (or row) of A are zero, then $\det A = 0$.

Theorem 1.3: If two columns (or rows) of A are proportional, then $\det A = 0$.

Theorem 1.4: If A is square of order n, $\det (cA) = c^n \det A$.

Theorem 1.5: Let A be a square matrix of order n. Then the system of homogeneous linear equations $A(X_1, X_2, \ldots, X_n) = AX = 0$ has a nontrivial solution if $\det A = 0$.

If we delete some rows and columns of a matrix A, the remaining elements form a submatrix of A. A square submatrix of A is called principal if its diagonal is part of the diagonal of A. The rank of A is the order of the largest square submatrix of A whose determinant is nonzero.

1.3 A STATEMENT OF THE CLASSIC PROBLEMS

As we mentioned previously, factor analysis is concerned with the study, or factoring, of a correlation matrix. Having discussed the

logical mode of representing the matrix and associated items of interest, let us restate the factor analysis problem.

Theorem 1.6: For every correlation matrix R there exists a corresponding factor matrix F such that

$$FF^T = R.$$

Furthermore,

Theorem 1.7: There exists an infinite number of factor matrices F which reproduce any given correlation matrix R.

The problem, then, is not only to determine F but to find an F which is most likely to satisfy a given set of initial conditions. A factor analysis is done in two stages:

Stage 1: Factoring problem--find an F such that $FF^T = R$ and also such that the column order of F is the minimal rank of R.

Stage 2: Rotation problem--rotate the arbitrary reference frame into a "preferred" or "simplifying" position.

In Stage 1 we mentioned the minimal rank of R. Ordinarily the rank of a matrix is fixed as soon as its elements are fixed. However, the diagonal elements of R have special meaning in that they represent the total variance of each variable. Due to the description of the model in terms of both common and unique factors, the total variance can be split into common factor variance (communality) and unique variance. The factor analysis of a correlation matrix with communalities on the diagonal, the reduced correlation matrix, will then yield only the common factor portion of the model. However the proportion of total variance ascribable to common factors is generally unknown. Thus, the matrix R is incomplete at the onset of a factor analysis. The communality problem consists of finding those diagonal elements of R that minimize the rank of R.

Once the rank of R has been established, the number of common factors is known and F can be determined by a variety of means (see Section V). By a previously stated theorem, however, there are an infinitude of F's which will do the job. The selection of the solution configuration--in other words, the relative number of high loadings per factor as well as degree of relationship among factors--is another classic problem. Probably the most commonly selected configuration is one called simple structure developed by L. L. Thurstone and offers the psychologist an optimal balance between statistical simplicity and psychological utility. There is little reason to believe that simple structure is of any real value outside the domain of a very special class of problems; however, intuitively it represents what may usually be desired in a factor solution (see Section V for detail).

Solution uniqueness is another classic problem which is important in defining the general usefulness of factor analysis. Assuming that a solution has been found which satisfies a given class of constraints and boundary conditions, what can we say about the uniqueness of this solution compared with a solution derived using another set of data points from the same multivariate population? Both solutions will be identical if infinite samples are used. However, from a practical viewpoint only a finite number of samples are possible and, in most cases, this number is small. Thus, the problem of uniqueness is really an error analysis of sampling effects on bivariate statistics and matrix operations.

Solution completeness is a problem which involves a decision to stop the factoring process after enough factors have been found (or extracted). This decision can be made in many ways depending on the kind of factor structure being derived. There does not exist a universal completeness criterion and the problem of completeness is often thought of as really the problem of communality selection.

In Stage 2 the rotation problem was stated as being one of finding a reference frame which provides a "preferred" or "simplified" position. The rotational aspects of factor analysis are the most

difficult to either understand or implement. This problem is by far the most important since the factor analyst has an infinitude of reference frames at his disposal from which he is to select one. Consider a similar problem whereby a point (x, y) in a plane is identified by its position relative to some orthogonal or non-orthogonal axes. The meaning of "preferred" or "simplified" is indeed vague and more or less has been defined by the analyst as a solution which fits closest to his hypothesized factor structure. In the case of a psychologist this factor structure has been characterized by Thurstone's simple structure. There are other structures which can be used, but they are not nearly developed to the extent of Thurstone's work.

1.4 METHOD OF APPROACH TO THE CLASSIC PROBLEMS

Naturally the classic problems--those problems which have defied analytical solutions--can not all be solved in one year of study. The very implication would be most insulting to the scientists who have spent lifetimes trying to clarify the intrinsic value of the methods. However, the time is ripe to establish a mathematical standard in factor analysis and provide mathematical explanations of the infinite solution space phenomenon as it effects uniqueness and other solution characteristics. The problems which have been considered in this study are the following:

1. communality estimate and completeness
2. uniqueness
3. rotation and interpretation.

The communality problem was approached from the standpoint of selecting diagonal elements which both minimized the rank of R and preserved the Gramian property (symmetric and the determinants of all principal submatrices are positive or zero). Several attempts were made using various iterative schemes and the technique of bordering has been found to solve the problem. Details of this technique are given in Section IV.

The uniqueness problem was approached from two angles--perturbing correlations and perturbing data. Correlations were bounded by

standard error intervals based on the sample size and effects noted on the factor structure. Data was randomized according to hypothetical correlations and distribution functions and new correlations derived and factored. Effects were again noted on the factor structure and empirical results are presented in Section IV.

The rotation and interpretation problem was approached through regression analysis in an attempt to provide a measure of importance for oblique or rotated factor loadings. The classic problem has been to identify or interpret factors using the factor loadings. Results are presented in Section V.

Section II

CORRELATION THEORY

2.1 INTRODUCTION

In this section we shall concern ourselves with the basic unit of factor analysis: the correlation coefficient. Factor analysis amounts to factoring a certain matrix, the correlation matrix, whose elements are the correlation coefficients. In this section we shall talk about these correlation coefficients. In 2.2 we shall generally define correlation and the coefficient describing it. In 2.3 we will consider different types of bivariate correlation coefficients and also compute examples. It will be seen that the coefficient most commonly used is Pearson's product-moment correlation coefficient. This coefficient will be interpreted geometrically in view of the factor model in 2.4. Its statistical significance and reliability is then discussed in 2.5. In 2.6 we consider how the product-moment correlation coefficient can be derived if data is missing from one variable or the other. In 2.7 we will touch briefly, for completeness, the areas of partial and multiple correlation coefficients.

2.2 DEFINITION OF CORRELATION AND THE COEFFICIENT

In factor analysis we are interested in the interrelationship of different variables, which we then analyze. But first we have to have a mathematical tool to express interrelationship between variables. This tool is given by the correlation coefficient.

Denote by X_j and X_k two variables each having values for N individuals. We first make the two variables comparable by deviating them, that is measuring their values from comparable zero points.

This is achieved by forming the <u>deviates</u>:

$$x_j = X_j - \overline{X}_j , \quad \overline{X}_j = \frac{1}{N} \sum_{i=1}^{N} X_{ji} = \text{the mean of variable } X_j$$

and

$$x_k = X_k - \overline{X}_k , \quad \overline{X}_k = \frac{1}{N} \sum_{i=1}^{N} X_{ki} = \text{the mean of variable } X_k .$$

Basically we assume that the relationship between variables x_j and x_k is linear, so that in plotting their paired values (x_{ji}, x_{ki}), $i = 1,\ldots,N$, in a coordinate system, with the zero point at the means of the two variables, we can ideally lay a straight line through these points. It will, though, obviously not always be the case that the points be on a straight line. Then we try to fit a straight line to the points. Expressing the points on the line by $\bar{x}_{ji}$, then the line can be described by

$$\bar{x}_{ji} = ax_{ki}, \quad i = 1,\ldots,N,$$

where a is called the slope of the line. The slope shows the relationship between $\bar{x}_{ji}$ and x_{ki}, $i = 1,\ldots,N$. If $a = 1$, $\bar{x}_{ji} = x_{ki}$ and the relationship is perfect; if $a = 0$, there does not exist any relationship between $\bar{x}_{ji}$ and x_{ki}, $i = 1,\ldots,N$. So we are interested in a which will later constitute our coefficient of correlation.

$\bar{x}_{ji} = ax_{ki}$ is a line "fitted" to the points. The condition for it is a "least square fit", that is,

$$\sum_{i=1}^{N} (x_{ji} - \bar{x}_{ji})^2 = \text{minimum.}$$

Then we have

$$\sum_{i=1}^{N} (x_{ji} - \bar{x}_{ji})^2 = \sum_{i=1}^{N} (x_{ji} - ax_{ki})^2$$

$$= \sum_{i=1}^{N} x_{ji}^2 - 2a \sum_{i=1}^{N} x_{ji}x_{ki} + a^2 \sum_{i=1}^{N} x_{ki}^2 ,$$

with the condition to choose a so that this expression is a minimum. Therefore we differentiate the expression with respect to a and set the result equal to zero, obtaining

$$-2 \sum_{i=1}^{N} x_{ji}x_{ki} + 2a \sum_{i=1}^{N} x_{ki}^2 = 0.$$

Therefore,

$$a = \frac{\sum_{i=1}^{N} x_{ji}x_{ki}}{\sum_{i=1}^{N} x_{ki}^2} .$$

This is the formula for the slope of a line fitting the trend of paired measures so as to minimize the x_j residuals. It is called the regression formula for x_j on x_k. Dividing numerator and denominator by N yields

$$a = \frac{\sum_{i=1}^{N} x_{ji}x_{ki}}{N \sum_{i=1}^{N} \frac{x_{ki}^2}{N}} .$$

We have

$$\sum_{i=1}^{N} \frac{x_{ki}^2}{N} = \sigma_k^2 = \text{the variance of } x_k.$$

Thus

$$a = \frac{\sum_{i=1}^{N} x_{ji}x_{ki}}{N\sigma_k^2} .$$

The slope a, however, is still greatly affected by the relative variabilities of the measures x_j and x_k. We make x_j and x_k comparable by dividing them each by their standard deviation. Calling $Z_j = x_j/\sigma_j$ and $Z_k = x_k/\sigma_k$ standardized variables, and naming the slope in this case r_{jk}, we obtain

$$\frac{\bar{x}_{ji}}{\sigma_j} = r_{jk}\frac{x_{ki}}{\sigma_k}, \quad \text{or} \quad \bar{x}_{ji} = \frac{r_{jk}\cdot x_{ki}\cdot\sigma_j}{\sigma_k}, \quad i = 1,\ldots,N.$$

From this follows

$$a = \frac{r_{jk}\cdot\sigma_j}{\sigma_k}, \quad \text{or} \quad r_{jk} = a\frac{\sigma_k}{\sigma_j},$$

and hence,

$$r_{jk} = \frac{\sum_{i=1}^{N} x_{ji}x_{ki}}{N\sigma_k^2}\cdot\frac{\sigma_k}{\sigma_j} = \frac{\sum_{i=1}^{N} x_{ji}x_{ki}}{N\sigma_j\sigma_k} = \frac{\sum_{i=1}^{N} Z_{ji}Z_{ki}}{N}.$$

r_{jk} is called Pearson's product-moment correlation coefficient between the standardized variables Z_j and Z_k.

2.3 TYPES OF CORRELATION COEFFICIENTS - BIVARIATE

In Section 2.2 we have derived Pearson's product-moment correlation coefficient. Besides this correlation coefficient there exist still other correlation coefficients, partly derivations from Pearson's r to take care of a specific nature of the variables.

In the present section we want to summarize most of the important correlation coefficients. We shall do this in a systematic way. So we shall define in A. Kendall's General Γ-correlation Coefficient, from which 1. Kendall's τ-, 2. Spearman's ρ-, and also 3. Pearson's r-correlation Coefficients can be derived as special cases.

Next we shall consider in 3. Correlation Coefficients for Dichotomized Variables (i.e. variables which are given by their frequencies in two classes). We shall discuss in 1. The Biserial Correlation Coefficient (a correlation coefficient for two variables, of which one is dichotomous and one has quantitative scores) in 2. The φ-coefficient (a correlation coefficient for two variables,

which are both truely dichotomous) and in 3. The Tetrachoric Correlation Coefficient (a coefficient for two variables which are both dichotomized from underlying normal and continuous distributions).

In C. we shall briefly consider Miscellaneous Correlation Coefficients by referring for the most part to some specific papers. These coefficients will be 1. The Contingency Coefficient, 2. Yule's Coefficient of Association and Yule's Coefficient of Colligation, and 3. Thorndike's Median Ratio Coefficient of Correlation.

Part D. then presents Examples to the aforementioned correlation coefficients.

Our discussion of all correlation coefficients will be very brief, mostly only a statement of the assumptions and of the basic definition. For standard error formulas and correction formulas one is referred to the references.

The answer to the question what correlation coefficient one should apply in a specific situation is given by the assumptions of the single coefficients, which are different for each coefficient.

A. Kendall's General Γ-correlation Coefficient

In the following we will consider the definition of the so-called Kendall's general Γ-correlation coefficient (Reference 3). We will state the necessary assumptions, the definition, and then we will derive three correlations coefficients from this general correlation coefficient, namely (1) Kendall's τ-correlation coefficient.

(2) Spearman's ρ-correlation coefficient.

(3) Pearson's product-moment correlation coefficient r.

Assumptions: A sample of N objects (subjects, individuals, observations, measurements) is considered relative to two properties (continuous variables) X and Y, exhibiting values $X_1,\ldots,X_N$ and $Y_1,\ldots,Y_N$ according to X and Y. To any pair of individuals i and j we will allot an X-score, denoted by a_{ij} and a Y-score b_{ij}, subject to $a_{ij} = -a_{ji}$, $b_{ij} = -b_{ji}$.

Definition 2.1: Kendall's general Γ-correlation coefficient is defined as

$$\Gamma = \frac{\sum_{i,j=1}^{N} a_{ij}b_{ij}}{\sqrt{\sum_{i,j=1}^{N} a_{ij}^2 \sum_{i,j=1}^{N} b_{ij}^2}}$$

with $a_{ij} = 0$ if $i = j$.

Now let us adopt three special methods of scoring and derive Kendall's τ-correlation coefficient, Spearman's ρ-correlation coefficient, and Pearson's product-moment correlation coefficient r.

1. Kendall's τ-correlation Coefficient

Assumptions: Suppose the values $X_1,\ldots,X_N$ are ranks, where we adopt the following definition for the term rank:

Definition 2.2: If N objects are arranged in order according to some property, which they all possess in a varying degree, the objects are said to be ranked. Each object has a rank, expressed as a natural number between 1 and N.

Denote them by $p_1,\ldots,p_N$. Correspondingly denote the ranks $Y_1,\ldots,Y_N$ by $q_1,\ldots,q_N$. Consider the pair of individuals i and j. Choose the following scores:

$$a_{ij} = 1, \text{ if } p_i < p_j \qquad b_{ij} = 1, \text{ if } p_i < p_j$$

and

$$a_{ij} = -1, \text{ if } p_i > p_j \qquad b_{ij} = -1, \text{ if } p_i > p_j.$$

Considering then the denominator in

$$\Gamma = \frac{\sum_{i,j=1}^{N} a_{ij}b_{ij}}{\sqrt{\sum_{i,j=1}^{N} a_{ij}^2 \sum_{i,j=1}^{N} b_{ij}^2}} ,$$

we observe

$$\sum_{i,j=1}^{N} a_{ij}^2 = \text{number of terms } a_{ij} = \frac{N!}{(N-2)!} = N(N-1)$$

and

$$\sum_{i,j=1}^{N} b_{ij}^2 = N(N-1).$$

Therefore we obtain

$$\text{denominator} = \sqrt{N(N-1)\,N(N-1)} = N(N-1).$$

Considering the numerator, we observe that

$$\sum_{i,j=1}^{N} a_{ij}b_{ij} = 2S ,$$

where S is the total score (sum of number of (+1)-scores and number of (-1)-scores), twice because any given pair (i, j) occurs once as (i, j) and once as (j, i) in the summation. We therefore obtain Kendall's τ-correlation coefficient, denoting Γ by τ, as

$$\tau = \frac{2S}{N(N-1)}$$

2. Spearman's ρ-correlation Coefficient

(Other names: Spearman's rank correlation, Spearman's rank difference method)

Assumption: Choose the following method of scoring:

$$a_{ij} = p_j - p_i$$
$$b_{ij} = q_j - q_i \qquad i,j = 1,\ldots,N .$$

We have

$$\sum_{i,j=1}^{N} a_{ij}^2 = \sum_{i,j=1}^{N} b_{ij}^2$$

or

$$\sum_{i,j=1}^{N} (p_j - p_i)^2 = \sum_{i,j=1}^{N} (q_j - q_i)^2 .$$

Then

$$r = \frac{\sum_{i,j=1}^{N} (p_j - p_i)(q_j - q_i)}{\sqrt{\sum_{i,j=1}^{N} (p_j - p_i)^2 \sum_{i,j=1}^{N} (q_j - q_i)^2}}$$

$$= \frac{\sum_{i,j=1}^{N} (p_j - p_i)(q_j - q_i)}{\sum_{i,j=1}^{N} (p_j - p_i)^2}$$

Considering the numerator,

$$\sum_{i=1}^{N} \sum_{j=1}^{N} (p_j - p_i)(q_j - q_i) = \sum_{i=1}^{N} \sum_{j=1}^{N} p_j q_j + \sum_{i=1}^{N} \sum_{j=1}^{N} p_i q_i$$

$$- \sum_{i=1}^{N} \sum_{j=1}^{N} (p_i q_j + p_j q_i)$$

$$= 2N \sum_{i=1}^{N} p_i q_i - 2 \sum_{i=1}^{N} p_i \sum_{j=1}^{N} q_j$$

$$= 2N \sum_{i=1}^{N} p_i q_i - 2 \left[\frac{N}{2}(1 + N)\right]^2$$

$$= 2N \sum_{i=1}^{N} p_i q_i - \frac{N^2}{2}(1 + N)^2.$$

Denote by $S(d)$ the sum of the differences $p_i - q_i$, $i = 1,\ldots,N$. Then

$$S(d^2) = \sum_{i=1}^{N} (p_i - q_i)^2 = \sum_{i=1}^{N} p_i^2 - 2 \sum_{i=1}^{N} p_i q_i + \sum_{i=1}^{N} q_i^2$$

$$= 2 \sum_{i=1}^{N} p_i^2 - 2 \sum_{i=1}^{N} p_i q_i ,$$

and therefore

$$\sum_{i=1}^{N} p_i q_i = \sum_{i=1}^{N} p_i^2 - \frac{S}{2}(d^2) .$$

We therefore obtain for the numerator

$$\sum_{i=1}^{N}\sum_{j=1}^{N}(p_j - p_i)(q_j - q_i) = 2N\sum_{i=1}^{N}p_i^2 - NS(d^2) - \frac{N^2}{2}(1+N)^2$$

$$= \frac{1}{6}N^2(N^2-1) - NS(d^2) .$$

Considering now the denominator

$$\sum_{i,j=1}^{N}(p_j - p_i)^2 = \sum_{i=1}^{N}\sum_{j=1}^{N}p_j^2 - 2\sum_{i=1}^{N}\sum_{j=1}^{N}p_jp_i + \sum_{i=1}^{N}\sum_{j=1}^{N}p_i^2$$

$$= 2N\sum_{i=1}^{N}p_i^2 - 2\sum_{i=1}^{N}p_i\sum_{j=1}^{N}p_j$$

$$= 2N\sum_{i=1}^{N}p_i^2 - 2\left(\sum_{i=1}^{N}p_i\right)^2$$

$$= 2N\left[\frac{1}{6}N(N+1)(2N+1)\right] - 2\left[\frac{N}{2}(1+N)\right]^2$$

$$= \frac{1}{6}N^2(N^2-1) .$$

Thus

$$\Gamma = \frac{\frac{1}{6}N^2(N^2-1) - NS(d^2)}{\frac{1}{6}N^2(N^2-1)} .$$

Denoting Γ by ρ, we obtain

$$\rho = 1 - \frac{6S(d^2)}{N(N^2-1)} .$$

3. Pearson's Product-moment Correlation Coefficient r

Assumptions: $a_{ij} = X_j - X_i$

$$b_{ij} = Y_j - Y_i .$$

Then

$$r = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(X_j - X_i)(Y_j - Y_i)}{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N}(X_j - X_i)^2 \sum_{i=1}^{N}\sum_{j=1}^{N}(Y_j - Y_i)^2}}$$

$$= \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}X_jY_j - \sum_{i=1}^{N}\sum_{j=1}^{N}X_jY_i - \sum_{i=1}^{N}\sum_{j=1}^{N}X_iY_j + \sum_{i=1}^{N}\sum_{j=1}^{N}X_iY_i}{\sqrt{\left(\sum_{i=1}^{N}\sum_{j=1}^{N}X_j^2 - 2\sum_{i=1}^{N}\sum_{j=1}^{N}X_jX_i + \sum_{i=1}^{N}\sum_{j=1}^{N}X_i^2\right)\left(\sum_{i=1}^{N}\sum_{j=1}^{N}(Y_j - Y_i)^2\right)}}$$

$$= \frac{2N\sum_{j=1}^{N}X_jY_j - 2\sum_{i=1}^{N}Y_i\sum_{j=1}^{N}X_j}{\sqrt{\left(2N\sum_{j=1}^{N}X_j^2 - 2\sum_{i=1}^{N}X_i\sum_{j=1}^{N}X_j\right)\left(\sum_{i=1}^{N}\sum_{j=1}^{N}(Y_j - Y_i)^2\right)}}$$

$$= \frac{2\left[N\sum_{j=1}^{N}X_jY_j - \sum_{i=1}^{N}Y_i\sum_{j=1}^{N}X_j\right]}{2\sqrt{\left(N\sum_{j=1}^{N}X_j^2 - \left(\sum_{i=1}^{N}X_i\right)^2\right)\left(N\sum_{j=1}^{N}Y_j^2 - \left(\sum_{i=1}^{N}Y_i\right)^2\right)}} .$$

Denoting r by r we obtain

$$r = \frac{\frac{\sum_{j=1}^{N}X_jY_j}{N} - \bar{Y}\bar{X}}{\sqrt{\left(\frac{\sum_{j=1}^{N}\bar{Y}_j^2}{N} - \bar{X}\right)\left(\frac{\sum_{j=1}^{N}\bar{Y}_j^2}{N} - \bar{Y}\right)}} ,$$

where

$$\frac{\sum_{i=1}^{N} X_i}{N} = \bar{X} \quad \text{and} \quad \frac{\sum_{i=1}^{N} Y_i}{N} = \bar{Y} .$$

The product-moment correlation coefficient is mostly considered with standardized variables, that is with zero mean and unit variance of the variables. To achieve this

$$\frac{\frac{\sum_{j=1}^{N} X_j Y_j}{N} - \bar{Y}\bar{X}}{\sqrt{\left(\frac{\sum_{j=1}^{N} \bar{X}_j^2}{N} - \bar{X}\right)\left(\frac{\sum_{j=1}^{N} \bar{Y}_j^2}{N} - \bar{Y}\right)}}$$

$$= \frac{\frac{\sum_{j=1}^{N} X_j Y_j}{N} - \bar{X}\frac{\sum_{j=1}^{N} Y_j}{N} - \bar{Y}\frac{\sum_{j=1}^{N} X_j}{N} + \frac{\sum_{j=1}^{N} \bar{X}\bar{Y}}{N}}{\sqrt{\left(\frac{\sum_{j=1}^{N} X_j^2}{N} - 2\frac{\sum_{j=1}^{N} X_j \bar{X}}{N} + \bar{X}^2\right)\left(\frac{\sum_{j=1}^{N} Y_j^2}{N} - 2\frac{\sum_{j=1}^{N} Y_j \bar{Y}}{N} + \bar{Y}^2\right)}}$$

$$= \frac{\frac{\sum_{j=1}^{N} (X_j - \bar{X})(Y_j - \bar{Y})}{N}}{\sqrt{\left(\frac{\sum_{j=1}^{N} (X_j - \bar{X})^2}{N}\right)\left(\frac{\sum_{j=1}^{N} (Y_j - \bar{Y})^2}{N}\right)}} .$$

Let

$$x_j = X_j - \bar{X} ,$$

$$y_j = Y_j - \bar{Y} ,$$

$$\sigma_{X_j} = \sqrt{\frac{\sum_{j=1}^{N} x_j^2}{N}} ,$$

$$\sigma_{Y_j} = \sqrt{\frac{\sum_{j=1}^{N} y_j^2}{N}} ,$$

$$Z_j = \frac{x_j}{\sigma_{X_j}} ,$$

and

$$Z'_j = \frac{y_j}{\sigma_{Y_j}} .$$

Then

$$\frac{\frac{\sum_{j=1}^{N} X_j Y_j}{N} - \overline{Y}\,\overline{X}}{\sqrt{\left(\frac{\sum_{j=1}^{N} X_j^2}{N} - \overline{X}\right)\left(\frac{\sum_{j=1}^{N} Y_j^2}{N} - \overline{Y}\right)}} = \frac{\sum_{j=1}^{N} x_j y_j}{N\sqrt{\left(\frac{\sum_{j=1}^{N} x_j^2}{N}\right)\left(\frac{\sum_{j=1}^{N} y_j^2}{N}\right)}}$$

$$= \frac{\sum_{j=1}^{N} x_j y_j}{N\sigma_{X_j}\sigma_{Y_j}}$$

$$= \frac{\sum_{j=1}^{N} Z_j Z'_j}{N} .$$

Remarks: Depending on the method of scoring the difference between the observations i and j for one variable, one obtains from Kendall's general Γ-correlation coefficient the τ-, ρ-, and r-coefficient.

The scoring for τ is therein the simplest one, assigning a 1 or -1 to this difference, thus not looking at all on how far apart the two observations are. The scoring for ρ is more involved, taking into account the actual difference of the observations by way of their ranking difference. For this reason ρ can be considered as the product-moment correlation coefficient between ranks. Scoring for r takes into account all the information by way of the actual difference between the measurements.

The choice of either one of the coefficients will depend on the data available. If actual measurements for continuous variables are available r is preferable to ρ and τ. If only data in the form of ranks are available, ρ is preferable to τ.

Pearson's product-moment correlation coefficient is the most important correlation coefficient for factor analysis, since its assumptions--rectilinearity and continuity of the variables, made for the derivation of this coefficient--are the ones which are mostly fulfilled by the variables involved in factor analysis.

B. Correlation Coefficients for Dichotomized Variables

1. The Biserial Correlation Coefficient

Assumptions: Let X_j and X_k be two variables. Consider one of them, say X_j, as dichotomous (or being reduced to dichotomy) under the assumption though, that it is really continuous, while we have only categorical information. Assume further that the dichotomized variable has a normal distribution, that the whole sample distribution is present, and that the two tails of the distribution fit together into a whole normal distribution. Looking only upon the two tails would make the coefficient, which will now be defined, too high. Consider the second variable X_k as having quantitative scores, no assumption made about its distribution. Assume a sample size of at least 50.

Denote the two categories of X_j by X_{j1} and X_{j2}. Let N be the total number of individuals, the sum of the number of individuals N_1 for X_{j1} and of the number of individuals N_2 for X_{j2}.

Definition 2.3: The biserial correlation coefficient is defined as

$$r = \left(\frac{M_p - M_t}{\sigma_t}\right)\left(\frac{p}{Y}\right)$$

where the following notation is adopted (Reference 4):

M_p = the meanscore on X_j of the individuals in category X_{j1} or X_{j2}, whichever is the larger

M_t = the meanscore on X_j of the individuals in X_{j1} and X_{j2} together

σ_t = the standard deviation of X_j for the entire distribution

p = the proportion $\frac{N_1}{N}$ or $\frac{N_2}{N}$, whichever is corresponding to the category with the higher mean on X_j

Y = the ordinate at the point of truncation of the normal distribution

Remarks:

a. If the dichotomized variable cannot be assumed to be continuous and normally distributed, Richardson and Stalnaker (Reference 5) suggest another form of the biserial correlation coefficient.

b. If one wants to look only upon the two tails of the distribution, which is often wanted in educational and sociological research, in other words, if one wants to look upon so-called "widespread classes", Peters and Van Voorhis (Reference 6) suggest a "biserial correlation coefficient from widespread classes".

c. Pearson (Reference 7) suggests a coefficient, called biserial eta, based on the assumption that one variable is given by alternative and the other by multiple categories.

2. The ϕ-coefficient

(Other name: Four-point Coefficient)

Assumptions: The two variables under consideration have to be truly dichotomcus. Let X_j and X_k be two variables with categories X_{j1}, X_{j2} and X_{k1}, X_{k2} respectively. Then establish the following table of frequencies a, b, c, d. Let the four cells be consistent with the quadrants of a coordinate system, represented by the signs.

X_k (2×2-fold table)

X_j	$-X_{k1}$	$+X_{k2}$	
$+X_{j1}$	a	b	a + b
$-X_{j2}$	c	d	c + d
	a + c	b + d	a + b + c + d = N

N = number of measurements, observations
= total frequency

Example

X_j = employment	X_{k1} = women	X_{k2} = men	
X_{j1} = employed	a	b	a + b
X_{j2} = unemployed	c	d	c + d
	a + c	b + d	N

X_k = sex

e.g. a = number of employed women
b = number of employed men
N = number of women and men, employed or unemployed.

Definition 2.4: The ϕ-coefficient is defined as

$$\phi = \frac{(bc) - (ad)}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} .$$

a. If we assume X_j and X_k to be dichotomous, while they are actually continuous, the ϕ-coefficient is to be considered as an estimate of Pearson's r. In order to obtain a good estimate of r a table (Reference 8) is available which gives a value k, by which ϕ has to be divided.

In general ϕ divided by k corresponds very closely to tetrachoric r (the correlation coefficient which is customarily applied to dichotomized, but really continuous data). So, if computing diagrams for tetrachoric r are not available, $\frac{\phi}{k}$ might be the most approximate measure for tetrachoric r.

b. In order to cut out the influence of extreme values, which go into the computation of the ϕ-coefficient, originating from extreme cuts in the distribution, ϕ is better divided by the maximum possible value consistent with the given marginal values, ϕ max. ϕ divided by ϕ max is probably the best correlation coefficient in use for dichotomized variables.*

3. The Tetrachoric Correlation Coefficient

Assumptions: Let X_j and X_k be the two variables under consideration. Assume that the data for both variables are in terms of dichotomies, but that both variables are really continuous and normal in distribution.

Definition of the Coefficient: The statistical considerations necessary for the derivation of the tetrachoric correlation coefficient are extensive. We will state here two of the formulas, used to compute the coefficient.

Again denote the two categories of X_j by X_{j1}, X_{j2} and the categories of X_k by X_{k1}, X_{k2}. a, b, c, d are notations for frequencies.

* E. E. Cureton. Note on ϕ/ϕ max. Psychometrika, 1959, 24, p.89.

		X_k	
		$-X_{k1}$	$+X_{k2}$
X_j	$+X_{j1}$	a	b
	$-X_{j2}$	c	d

The statistical derivation terminates in a formula involving double integration, which can be solved for the tetrachoric correlation coefficient r, yielding a very complicated formula for r (Reference 9).

In putting the restriction upon the problem of cutting the distributions at the mean, the following formula for the tetrachoric correlation coefficient can be arrived at:

$$r = \sin 2\pi \frac{(bc - ad)}{N^2}, \quad N = a + b + c + d.$$

The assumption of equal dichotomies might be a crude one for certain problems. So, Pearson develops (Reference 9) empirical formulas that give approximately correct r's, the mean error in 15 trials being less than 4 per cent. The simplest of these approximate formulas is the following one

$$r = \cos\left(\pi \frac{\sqrt{ad}}{\sqrt{ad} + \sqrt{bc}}\right)$$

where no restriction is put on the point of dichotomy. H. W. Eber (Reference 10) uses this formula for computing a correlation matrix for 3,000 variables.

Remarks

a. In order to facilitate the labor involved in computing tetrachoric correlation coefficients, Chesire, Saffir, and Thurstone (Reference 11) prepared a set of computing diagrams. These diagrams

are advisable to be used whenever the coefficient is not required to be of high accuracy. Other diagrams are designed by Hamilton (Reference 12).

b. As for the biserial correlation coefficient Peters and Von Voorhis (Reference 6) develop a tetrachoric correlation coefficient from widespread classes.

c. The tetrachoric correlation coefficient is one of the coefficients for factor analysis more often used besides the product-moment coefficients. To use this coefficient is thereby a necessary condition, if the data are reported in dichotomies only. If the dichotomies are derived, though, by cutting continuous data at some point, it should be strongly considered to employ product-moment coefficients instead, since the tetrachoric correlation technique loses some of the available information.

C. Miscellaneous Correlation Coefficients

1. The Contingency Coefficient

The contingency coefficient is applied when variables X_j and X_k both can be classified into two or more categories, and when these categories are not quantitative but qualitative. The formula of the contingency coefficient makes use of the chi-square statistic.

Definition 2.5: The contingency coefficient is defined as

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

Under certain conditions C is equivalent to Pearson's product-moment correlation coefficient. If the variables are continuous, correction formulas exist, see References 13, 14, and 15.

2. Yule's Coefficient of Association and Yule's Coefficient of Colligation

In connection with the ϕ-correlation coefficient Yule (Reference 16) considers two correlation coefficients, based on a four-fold table:

Definition 2.6: Yule's coefficient of association is defined as

$$Q = \frac{bc - ad}{ad + bc}$$

Definition 2.7: Yule's coefficient of colligation is defined as

$$\omega = \frac{\sqrt{bc} - \sqrt{ad}}{\sqrt{ad} + \sqrt{bc}}$$

The coefficient is equal to ϕ, if the four-fold table is "equalized", that is

$\sqrt{ad}$	$\sqrt{bc}$
$\sqrt{bc}$	$\sqrt{ad}$

3. Thorndike's Median Ratio Coefficient of Correlation

Thorndike (References 17 and 18) developed a correlation coefficient, which, under certain conditions (Kelley, Reference 13) is equal to the product-moment correlation coefficient.

Let the variables x_j and x_k be deviates from the mean and let σ_j and σ_k be corresponding standard deviations. Supposing the relation of variables x_j and x_k to be rectilinear the coefficient of correlation, defined as follows, represents an inference about the general drift of the relation.

Definition 2.8:
Thorndike's median ratio coefficient of correlation is defined as

r = median of the 2N ratios

$$\frac{x_j i/\sigma_j}{x_{ki}/\sigma_k}$$

and

$$\frac{x_{ki}/\sigma_k}{x_{ji}/\sigma_j} \quad , i=1,\ldots, N.$$

For simplifying computing formulas see Thorndike (Reference 17). See also Kelley (Reference 13).

D. Examples

1. Example for Kendall's τ-correlation Coefficient

The coefficient is

$$\tau = \frac{S}{\frac{N}{2}(N-1)}$$

Before computing the coefficient on the basis of the data in Table 1, we have to consider briefly how to take care of ties in the data. Let $t(u)$ be the number of equally ranked individuals, then there are $\frac{1}{2}t(t-1)$ pairs to take care of.

Denote by $$T = \frac{1}{2}\sum_t t(t-1)$$

$$U = \frac{1}{2}\sum_t u(u-1),$$

where $\sum_t$ means summation over all sets of ties.
Then τ is computed as

$$\tau = \frac{S}{\sqrt{\frac{1}{2}N(N-1) - T}\ \sqrt{\frac{1}{2}N(N-1) - U}}$$

This is the appropriate form of τ if ties arise in the data. The computation will be clear from the example. The formula is stated and discussed by Kendall (Reference 3).

Table 1

Measurements for Two Variables--Weight and Height on 10 Individuals*

Individual	Ranks	Weight = X_1	Ranks	Height = X_2	d = rank difference	d^2
A	6	165.00	8	177.80	-2	4
B	1	189.50	1	187.60	0	0
C	10	128.00	10	169.00	0	0
D	9	144.00	4.5**	181.50	4.5	20.25
E	7	156.50	7	179.70	0	0
F	8	145.50	9	172.90	-1	1
G	5	166.00	4.5**	181.50	0.5	0.25
H	3	178.00	2	185.30	1	1
I	2	182.50	6	181.00	-4	16
J	4	167.50	3	182.35	1	1

* The measurements for the 10 individuals were picked randomly from a set of measurements for 130 individuals.

** Individuals D and G are tied for ranks 4 and 5. It is common use to rank each individual by the average of the tied ranks.

To find S we have to compare each individual i with each individual j. We score

$$a_{ij} = 1, \text{ if } p_i < p_j$$

$$a_{ij} = -1, \text{ if } p_i > p_j$$

$$a_{ij} = 0, \text{ if } p_i = p_j \ .$$

Listing the results, also for b_{ij}, we obtain:

	Individual i Compared with Individual j i,j = A,B,...,J for X_1	Scores Multiplied	Individual i Compared with individual j where i,j = A,B,...,J for X_2
A with B	-1	1	-1
C	1	1	1
D	1	-1	-1
E	1	-1	-1
F	1	1	1
G	-1	1	-1
H	-1	1	-1
I	-1	1	-1
J	-1	1	-1
B with C	1	1	1
D	1	1	1
E	1	1	1
F	1	1	1
G	1	1	1
H	1	1	1
I	1	1	1
J	1	1	1
C with D	-1	1	-1
E	-1	1	-1
F	-1	1	-1
G	-1	1	-1
H	-1	1	-1
I	-1	1	-1
J	-1	1	-1

	Individual i Compared with Individual j i,j = A,B,...,J for X_1	Scores Multiplied	Individual i Compared with individual j where i,j = A,B,...,J for X_2
D with E	1	-1	-1
F	-1	-1	1
G	-1	0	0
H	-1	1	-1
I	-1	1	-1
J	-1	-1	1
E with F	1	1	1
G	-1	1	-1
H	-1	1	-1
I	-1	1	-1
J	-1	1	-1
F with G	-1	1	-1
H	-1	1	-1
I	-1	1	-1
J	-1	1	-1
G with H	-1	1	-1
I	-1	-1	1
J	-1	1	-1
H with I	-1	-1	1
J	1	1	1
I with J	1	-1	-1

Then

S = (Sum of (+1)-scores) - (sum of (-1)-scores)

S = 36 - 8 = 28

It is

T = 0, there are no ties in X_1

$U = \frac{1}{2}(2 \cdot 1) = 1$, there is one tie in X_2.

Then we obtain

$$\tau = \frac{28}{\sqrt{\frac{1}{2}10(10-1)-0}\ \sqrt{\frac{1}{2}10(10-1)-1}}$$

$$= \frac{28}{\sqrt{45 \cdot 44}}$$

$$= 0.629$$

2. Example for Spearman's ρ-Correlation Coefficient

The coefficient is

$$\rho = 1 - \frac{6S(d^2)}{N(N^2-1)}$$

Also in Spearman's ρ-correlation coefficient we have to take care of ties in the ranking of the two correlated variables. Typify the ties by t and u and define

$$T = \frac{1}{12}\sum_t t(t^2-t)$$

$$U = \frac{1}{12}\sum_u u(u^2-u) \quad .$$

Then Kendall obtains two equations, deducing them from the general Γ-coefficient:

$$\rho = 1 - \frac{6(S(d)^2 + T + U)}{N(N^2 - 1)} \qquad (1)$$

or

$$\rho = \frac{\frac{1}{6}N(N^2-1) - S(d^2) - T - U}{\sqrt{\left[\frac{1}{6}N(N^2-1)-2T\right]\left[\frac{1}{6}N(N^2-1)-2U\right]}} \quad . \qquad (2)$$

For our data of Table 1 we obtain

$$T = 0$$
$$U = \frac{1}{12}\,2(2^2-2) = \frac{1}{3} \quad .$$

Then for Equation 1

$$\rho = 1 - \frac{6\left\{43.5 + \frac{1}{3}\right\}}{990}$$

$$= 0.734 .$$

Equation 2 yields

$$\rho = \frac{\frac{1}{6}(990) - 43.5 - \frac{1}{3}}{\sqrt{\left[\frac{1}{6}(990) - 0\right]\left[\frac{1}{6}(990) - \frac{2}{3}\right]}}$$

$$= 0.736$$

3. Example for Pearson's Product-Moment Correlation Coefficient r

We shall use the following form of the coefficient

$$r = \frac{\frac{\sum_{j=1}^{N} X_j Y_j}{N} - \overline{X}\ \overline{Y}}{\sqrt{\left(\sum_{j=1}^{N} \frac{X_j^2}{N} - \overline{X^2}\right)\left(\sum_{j=1}^{N} \frac{Y_j^2}{N} - \overline{Y^2}\right)}}$$

We again will use the data from Table 1:

Therefore

$$X = X_1 , \quad Y = X_2 , \quad N = 10 .$$

Then one computes

$$\sum_{j=1}^{10} \frac{X_{1j} X_{2j}}{10} = 29262.37$$

$$\bar{X}_1 = \sum_{j=1}^{10} \frac{X_{1j}}{10} = 162.25, \quad \bar{X}_1^2 = 26325.06$$

$$\bar{X}_2 = \sum_{j=1}^{10} \frac{X_{2j}}{10} = 179.87, \quad \bar{X}_2^2 = 32353.21$$

$$\sum_{j=1}^{10} \frac{X_{1j}^2}{10} = 26652.02$$

$$\sum_{j=1}^{10} \frac{X_{2j}^2}{10} = 32378.72$$

and one obtains by substituting these values in the formula for r.

$$r = 0.859$$

4. Example for the Biserial Correlation Coefficient

$$r = \left(\frac{M_p - M_t}{\sigma_t}\right)\left(\frac{P}{Y}\right)$$

We shall use the coefficient to determine the relation between the variable X_1 = size of family and variable X_2 = tendency of children to leave school before the age of eighteen. The data of X_2 are given by the two categories: X_{21} = children, who remained in school according to the size of family, X_{22} = children who left school according to the size of family. The data are laid out in Table 2.

Table 2

Data for an Example of the Biserial Correlation Coefficient

Size of Family X_1 (Class Marks X_j)	Children remaining in School X_{21} (Frequencies f_{1j})	Children left school X_{22} (Frequencies f_{2j})	Total ($f_{1j} + f_{2j}$)
12	2		2
11	4	3	7
10	4	2	6
9	4	8	12
8	20	3	23
7	10	17	27
6	24	12	36
5	18	18	36
4	30	10	40
3	34	12	46
2	34	10	44
1	16	5	21
	200	100	300

The example is from Reference 6 and the measures are from Reference 19.

Let us first compute the meanscore on X_2 in categories X_{21}, X_{22}, and $X_{21} + X_{22}$. Using the mean formula for grouped data:

$$\overline{X}_{21} = \sum_{j=1}^{12} \frac{f_{1j}X_j}{200} = \frac{2 \cdot 12 + 4 \cdot 11 + \cdots + 16 \cdot 1}{200} = \frac{914}{200} = 4.57$$

$$\overline{X}_{22} = \sum_{j=1}^{12} \frac{f_{2j}X_j}{100} = \frac{3 \cdot 11 + 2 \cdot 10 + \cdots + 5 \cdot 1}{100} = \frac{531}{100} = 5.31$$

$$\overline{X_{21} + X_{22}} = \sum_{j=1}^{12} \frac{(f_{1j} + f_{2j})X_j}{300} = \frac{2 \cdot 12 + 7 \cdot 11 + \cdots + 21 \cdot 1}{300} = \frac{1445}{300} = 4.82$$

Such that

$$M_p = 5.31$$

$$M_t = 4.82$$

Now we compute the standard deviation of X_2 for $X_{21} + X_{22}$. We use again the formula for grouped data:

$$\sigma_t = \sqrt{\frac{\sum_{j=1}^{12} (f_{1j} + f_{2j})(X_j - \overline{X_{21} + X_{22}})^2}{300}}$$

$$= \sqrt{\frac{2(12 - 4.82)^2 + 7(11 - 4.82)^2 + \ldots + 21(1 - 4.82)^2}{300}}$$

$$= 2.57$$

and

$$p = \frac{N_2}{N} = \frac{100}{300} = 0.33 \quad .$$

Then $y = 0.3635$, as taken from a table by Peters and Van Voorhis (Reference 6).

We now compute r as:

$$r = \left(\frac{5.31 - 4.82}{2.57}\right)\left(\frac{0.333}{0.3635}\right) = 0.175$$

5. Example for the ϕ-coefficient

The ϕ-coefficient is computed from a four-fold tables as are the tetrachoric correlation coefficient and Yule's two coefficients, the contingency coefficient can be computed from a four-fold or a manifold table. We will now demonstrate all above mentioned coefficients from an identical four-fold table, which will only be interpreted differently for the single coefficient under consideration, in order to allow for the special assumptions of this coefficient.

The ϕ-coefficient

$$\phi = \frac{bc - ad}{\sqrt{(a + b)\,(c + d)\,(a + c)\,(b + d)}}$$

we will lay out the data in Table 3.

Table 3. Data for an Example of the ϕ-coefficient

		X_2: X_{21}	X_{22}	
X_1: X_{11}		a = 665	b = 849	a + b = 1514
X_{12}		c = 1281	d = 205	c + d = 1486
		a + c = 1946	b + d = 1054	a + b + c + d = N = 3000

An interpretation is given: We want to determine the relationship of employment status (X_1) and sex classification (X_2), where both X_1 and X_2 are given by two categories X_{11} = being employed, X_{12} = being unemployed, X_{21} = women, X_{22} men. So, e.g., a = 665 represents the number of women, questioned in a sample of 3000 men and women (1946 women, 1054 men), who were employed out of the 1946 women in the sample. Note, that both variables, sex and employment, are truely dichotomous.

We obtain

$$\phi = \frac{849 \cdot 1281 - 665 \cdot 205}{\sqrt{1514 \cdot 1486 \cdot 1054 \cdot 1946}}$$

$$= 0443 \ .$$

That is, the relation between sex classification and employment status is a positive one. That means, for the data under consideration being a man and being employed are positively related.

Connected to the ϕ-coefficient are Yule's two coefficients. Their computation from the data in Table 3 gives the following results:

Yule's Coefficient of Association:

$$Q = \frac{bc - ad}{ad + bc}$$

$$= \frac{849 \cdot 1281 - 665 \cdot 205}{849 \cdot 1281 + 665 \cdot 205}$$

$$= 0.776 \ .$$

Yule's Coefficient of Colligation:

$$\omega = \frac{\sqrt{bc} - \sqrt{ad}}{\sqrt{ad} + \sqrt{bc}} = \frac{\sqrt{849 \cdot 1281} - \sqrt{665 \cdot 205}}{\sqrt{849 \cdot 1281} + \sqrt{665 \cdot 205}}$$

$$= 0.477 \ .$$

6. Example for the Tetrachoric Correlation Coefficient

We will use the cosine-formula of the coefficient

$$r = \cos\left(\pi \frac{\sqrt{ad}}{\sqrt{ad} + \sqrt{bc}}\right)$$

and apply it to the data of Table 3. As an interpretation of the fourfold table, consider the case that we have a sample of 3000 teachers divided into successful and unsuccessful ones and that we have information about how many of the successful and how many of the unsuccessful teachers have taken courses in pedagogy beyond 6 hours or less 6 hours. We want to know the relationship of teacher success and taking courses in pedagogy.

We set

X_1 = teacher success--X_{11} = successful

X_{12} = unsuccessful

X_2 = courses in pedagogy--X_{21} = beyond 6 hours

X_{22} = less than 6 hours.

Note, that one can think of both variables as being continuous, though they are represented as dichotomous.

$$r = \cos\left(\pi \frac{\sqrt{665 \cdot 205}}{\sqrt{665 \cdot 205} + \sqrt{849 \cdot 1281}}\right)$$

$$= 0.6811$$

The tetrachoric correlation coefficient computed from the tables of Pearson and his students has the value $r = 0.6633$ for the above data

Chesire, Saffir and Thurstone compute a value of $r = 0.6638$ for the considered data by their computing diagrams.

7. Example for the Contingency Coefficient

The coefficient

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}}$$

could be applied to determine relationship between two variables, each described in more than two categories. Let the variables, for example, be eye color of fathers (X_1) and eye color of sons (X_2). Each variable may be divided in many categories: X_{11} and X_{21} = brown X_{12} and X_{22} = grey, X_{13} and X_{23} = blue and so forth. We will apply the coefficient, though, to data reported in a four-fold table, thereby assuming that each variable above has only two categories. We will use the data reported in Table 3.

We will make use of a simple computing formula for χ^2 for the case of a four-fold table:

$$\chi^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} ,$$

proved for example in Reference 20.

We obtain

$$C = \sqrt{\frac{\frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}}{\left(N + \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}\right)}}$$

$$= \sqrt{\frac{N(ad-bc)^2}{N(a+b)(c+d)(a+c)(b+d) + N(ad-bc)^2}}$$

$$= \frac{-849 \cdot 1281 + 665 \cdot 205}{\sqrt{1514 \cdot 1486 \cdot 1054 \cdot 1946}} = -0.405 .$$

8. Example for Thorndike's Median Ratio Coefficient of Correlation

r = median of the $2N$ ratios

$$\frac{x_{ji}/\sigma_j}{x_{ki}/\sigma_k}$$

and

$$\frac{x_{ki}/\sigma_k}{x_{ji}/\sigma_j} \quad , \; i = 1,\ldots,N \quad .$$

We want to apply the coefficient to the data of Table 1.

Then

x_1 = weight
x_2 = height, and $N = 10$

We first have to set x_{1i}, x_{2i}, $i = 1,\ldots,10$, as deviates from the respective means. The means are $\bar{x}_1 = 162.25$ and $\bar{x}_2 = 179.87$. Subtracting $\bar{x}_1$ from all x_{1i}, $i = 1,\ldots,10$ and $\bar{x}_2$ from x_{2i}, $i = 1,\ldots,10$, we can compute the standard deviations as

$$\sigma_1 = \sqrt{\sum_{i=1}^{10} \frac{x_{1i}^2}{10}} \quad , \qquad \sigma_2 = \sqrt{\sum_{i=1}^{10} \frac{x_{2i}^2}{10}}$$

to get

$$\sigma_1 = 18.08$$

$$\sigma_2 = 5.22 \quad .$$

The ratios $y = x_{1i}/\sigma_1$ and $z = x_{2i}/\sigma_2$, i-= 1,...,10, are formed next. And then y/z and z/y are considered. Their median furnishes Thorndike's correlation coefficient. It is computed as r = 0.872 from the data of Table 1. Note its rather good agreement with Pearson's product moment r = 0.859.

2.4 GEOMETRIC INTERPRETATION OF PEARSON'S PRODUCT-MOMENT CORRELATION COEFFICIENT

In this section we want to interpret the correlation coefficient (if we talk about the correlation coefficient, we mean Pearson's product-moment correlation coefficient) in view of its geometric aspects with respect to the factor model.

Let us then assume an N-dimensional Euclidean space with a rectangular Cartesian Coordinate System, whose origin is denoted by $0:(0,\ldots,0)$ and whose unit points are denoted by $E_1:(1,0,\ldots,0),\ldots,E_N:(0,\ldots,0,1)$. Let us interpret the n variables Z_j as points represented in this system, the points and their coordinates denoted by $Z_j:(Z_{j1},\ldots,Z_{jN}) = Z_j:(Z_{ji})$. Such a representation for each of the n variables can be called a vector representation, each Z_j being named a vector. Let, further on, the N lines OZ_{ji} (i=1,...,N), each passing through the origin and one of the unit points be called coordinate axes.

Now let us make the following definitions:

Definition 2.9: For any two points $Z_j:(Z_{j1},\ldots,Z_{jN})$ and $Z_k:(Z_{k1},\ldots,Z_{kN})$ their distance is defined by

$$D(Z_jZ_k) = \sqrt{\sum_{i=1}^{N} (Z_{ji} - Z_{ki})^2} \quad .$$

If the distance of a point Z_j from the zero point is considered, it is called the norm denoted by

$$D(OZ_j) = D(Z_j) = \sqrt{\sum_{i=1}^{N} Z_{ji}^2}$$

Definition 2.10: Let the norm $D(OZ_j)$ be denoted by ρ_j. Then the angles, which the line OZ_j makes with the axes, denoted by $\theta_{ji} = \sphericalangle Z_j O Z_{ji}$, are called the direction angles of the line and their cosines are called the direction cosines, denoted by $\lambda_{ji} = \cos \theta_{ji}$. From the definition it follows that $\lambda_{ji} = \cos \theta_{ji} = Z_{ji}/\rho_j, i=1,\ldots,N$.

Now the following interesting implications can be made.

a. From

$$\rho_j = D(OZ_j) = \sqrt{\sum_{i=1}^{N} Z_{ji}^2}$$

follows

$$\rho_j^2 = \sum_{i=1}^{N} Z_{ji}^2 .$$

Since

$$\lambda_{ji}^2 = \cos^2 \theta_{ji} = \frac{Z_{ji}^2}{\rho_j^2} \quad , i=1,\ldots,N$$

it follows

$$\sum_{i=1}^{N} \lambda_{ji}^2 = \sum_{i=1}^{N} \cos^2 \theta_{ji} = \sum_{i=1}^{N} \frac{Z_{ji}^2}{\rho_j^2} = \frac{\sum_{i=1}^{N} Z_{ji}^2}{\sum_{i=1}^{N} Z_{ji}^2} = 1 \ .$$

That is, the sum of the squares of the direction cosines of a line in N-space is equal to unity.

b. Next, denote by λ_{ji} and λ_{ki} the direction cosines of the vectors Z_j and Z_k.

Then

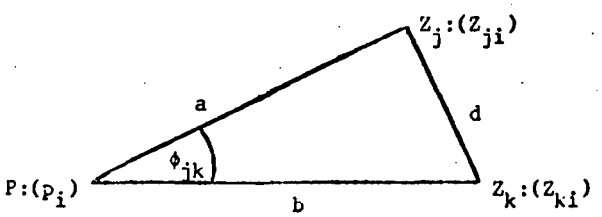
$$\lambda_{ji} = \frac{Z_{ji}}{\rho_j}$$

and

$$\lambda_{ki} = \frac{Z_{ki}}{\rho_k} \quad , \; i=1,\ldots,N.$$

We are now interested in the angle of separation of two lines in N-space, precisely said, in the cosine of this angle. We can derive an appropriate formula by using the direction cosine formulas and by referring to the trigonometric properties of a triangle in the plane, visualizing that, if two lines meet in a point, a plane can be drawn through the point containing the two lines. If the two lines do not meet in a point, a line can be drawn parallel to one of them, so that the line and the parallel form the angle we are interested in.

Denote the vertex of the angle by $P:(p_i)$, the angle by ϕ_{jk}, and distances as follows: $D(PZ_j) = a$, $D(PZ_k) = b$ and $D(Z_jZ_k) = d$. Then we can draw the following picture:



The coordinates of Z_j and Z_k can be expressed

as

$$Z_{ji} = p_i + a\lambda_{ji}, \; i=1,\ldots,N,$$

$$Z_{ki} = p_i + b\lambda_{ki}, \; i=1,\ldots,N.$$

Then, applying the law of cosines, we obtain

$$d^2 = a^2 + b^2 - 2ab \cos \phi_{jk}. \tag{3}$$

Applying the distance formula we obtain

$$d^2 = (D(Z_j Z_K))^2 = \sum_{i=1}^{N} (Z_{ji} - Z_{ki})^2$$

$$= \sum_{i=1}^{N} [(p_i + a\lambda_{ji}) - (p_i + b\lambda_{ki})]^2$$

$$= a^2 \sum_{i=1}^{N} \lambda_{ji}^2 + b^2 \sum_{i=1}^{N} \lambda_{ki}^2 - 2ab \sum_{i=1}^{N} \lambda_{ji}\lambda_{ki}$$

$$= a^2 + b^2 - 2ab \sum_{i=1}^{N} \lambda_{ji}\lambda_{ki} \quad . \tag{4}$$

This implies by identification of terms in Equations 3 and 4:

$$\cos \phi_{jk} = \sum_{i=1}^{N} \lambda_{ji}\lambda_{ki} \quad .$$

That is, the cosine of the angle of separation of two lines is given by the inner product of corresponding direction cosine vectors $(\lambda_{j1},\ldots,\lambda_{jN})$ and $(\lambda_{k1},\ldots,\lambda_{kN})$.

c. Since

$$\lambda_{ji} = \frac{Z_{ji}}{\rho_j} \quad , i=1,\ldots,N$$

$$\lambda_{ki} = \frac{Z_{ki}}{\rho_k} \quad , i=1,\ldots,N \quad ,$$

we obtain

$$\cos \phi_{jk} = \sum_{i=1}^{N} \lambda_{ji}\lambda_{ki} = \sum_{i=1}^{N} \frac{Z_{ji}Z_{ki}}{\rho_j \rho_k} .$$

It is

$$\rho_j = \sqrt{\sum_{i=1}^{N} Z_{ji}^2}$$

$$= \sqrt{N \frac{\sum_{i=1}^{N} Z_{ji}^2}{N}}$$

$$= \sqrt{N} ,$$

since

$$\sqrt{\frac{\sum_{i=1}^{N} Z_{ji}^2}{N}} = \text{standard deviation}$$

$$= 1, \text{ for standardized variables } Z_j .$$

Thus

$$\cos \phi_{jk} = \sum_{i=1}^{N} \frac{Z_{ji}Z_{ki}}{\rho_j \rho_k} = \sum_{i=1}^{N} \frac{Z_{ji}Z_{ki}}{N} = r_{jk} , \; j, k=1,\ldots,n.$$

These considerations yield the result, that the coefficient of correlation between two standardized variables is the cosine of the angle between their vectors in N-space.

d. Our geometric interpretation of a correlation coefficient, so far, started with the consideration of the given raw data, namely the n points $Z_j:(Z_{ji})$ in N-space. Then the cosine of the angle between two such vectors in N-space constitutes the coefficient of correlation between two variables.

Now we assume a correlation matrix R, computed from the raw data, and a factorization of this correlation matrix. According to the mathematical model underlying factor analysis each variable Z_j is now expressible by

$$Z_j' = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jm}F_m + a_jU_j,\ j=1,\ldots,n,$$

where the loadings $a_{ji}(i=1,\ldots,m)$ were obtained from the factorization. In this representation the n vectors Z_j are considered in the space of m common factors and n unique factors, the total-factor space. The vector representation of $Z_j(j=1,\ldots,n)$ in this space is denoted by $Z_j':(a_{j1}, a_{j2},\ldots, a_{jm}, 0,\ldots, 0, a_j, 0,\ldots, 0)$, the $a_{ji}(i=1,\ldots,m)$ denoting the coordinates of Z_j' with respect to the common-factor axes, the 0 and a_j denoting the coordinates of Z_j' with respect to the unique-factor axes. We now assume, that the system of common- and unique-factor axes is rectangular, that is, all factors are mutually orthogonal. Then the angle of separation of two vectors Z_j' and Z_k', represented in this system, is, according to the formula, discussed in (b)

$$\cos \phi_{jk}' = \sum_{i=1}^{m+n} \lambda_{ji}'\lambda_{ki}' = \sum_{i=1}^{m} \frac{a_{ji}a_{ki}}{\rho_j\rho_k},\ j,k=1,\ldots,n.$$

Since

$$\rho_j = \sqrt{\sum_{i=1}^{m} a_{ji}^2 + a_j^2} = 1, \text{ for } j=1,\ldots,n,$$

(Since

$$\sum_{i=1}^{m} a_{ji}^2 + a_j^2 = \text{total variance})$$

it follows that

$$\cos \phi_{jk}' = \sum_{i=1}^{m} a_{ji}a_{ki} = r_{jk}',\ j,k=1,\ldots,n.$$

Thus we obtain that the reproduced correlation coefficient (from the pattern of loadings) of any two variables Z'_j, Z'_k is equal to the cosine of the angle between their vectors in the total-factor space.

e. Our last consideration informed us about how a correlation coefficient is described if the variables are assumed to be represented in the total-factor space.

Finally, in factor analysis one usually does not consider the total-factor space but the space of m common factors only, that is, one regards the n vectors contained in an n-dimensional space, determined by the m factors. To obtain this m-dimensional space one considers the orthogonal projections of the n vectors from the total-factor space into the common-factor space of m dimensions and defines these orthogonal projections to be the vectors representing the variables in this space, denoted by

$$Z''_j:(a_{j1}, a_{j2},\ldots,a_{jm}),\ j=1,\ldots,n.$$

We assume a rectangular coordinate system to be set up in the common-factor space.

Considering now the angle of separation of two vectors Z''_j and Z''_k, represented in this space, one obtains:

$$\cos\ \phi''_{jk} = \sum_{i=1}^{m} \lambda''_{ji}\lambda''_{ki} = \sum_{i=1}^{m} \frac{a_{ji}a_{ki}}{\rho_j\rho_k}\ ,\ j,\ k=1,\ldots,n,$$

where

$$\rho_j = \sqrt{\sum_{i=1}^{m} a_{ji}^2} = \sqrt{h_j^2} = h_j,\ j=1,\ldots,n$$

(with h_j^2 = communality),

so that

$$\cos\ \phi''_{jk} = \sum_{i=1}^{m} \frac{a_{ji}a_{ki}}{h_j h_k}$$

and from (d):

$$\cos \phi''_{jk} = \frac{r'_{jk}}{h_j h_k} = r''_{jk}, \; j, k=1,\ldots,n.$$

Hence the cosine of the angle between two vectors which represent variables in the common-factor space is equal to the reproduced correlation coefficient, divided by the product of the square roots of the communalities of these two variables. We may call the obtained correlation coefficients r''_{jk} "the correlation coefficient between Z''_j and Z''_k corrected for uniqueness", since only if the two variables do not have any unique variance would r''_{jk} be equal to the reproduced correlation coefficient r'_{jk}.

2.5 SIGNIFICANCE AND RELIABILITY OF PEARSON'S PRODUCT-MOMENT CORRELATION COEFFICIENT

A statistical consideration that can be made on a Pearson product--moment correlation coefficient r_{jk} is the determination of its statistical significance. Since statistical significance of r_{jk} is dependent on sample size N, the following considerations will give us some important information about this dependence which we shall utelize even more at a later stage.

Let us first briefly consider what is meant by statistical significance. In statistical considerations mostly only sample information is available, on the basis of which one tries to make decisions about the population, from which the sample was drawn. The decisions are called statistical decisions. In attempting to reach decisions, one then makes assumptions about the population involved. These assumptions, which may or may not be true, are called statistical hypothesis. They mostly are statements about the probability distribution of the population in question. If we assume a certain hypothesis to be true and then find that results observed in a random sample differ markedly from those, which we expected under the hypothesis on the basis of pure chance using sampling theory, we would say that the observed differences are significant. We would then reject the hypothesis. Procedures which make it possible to decide whether to accept or reject a hypothesis or to determine whether observed samples differ significantly from expected

results are called tests of hypothesis or tests of significance. When one tests a hypothesis, the maximum probability with which one is willing to risk the error of rejecting a hypothesis when it should be accepted is called the level of significance of the test. Usually a 5% level of significance is chosen, that means we are 95% confident, that we have made the right decision in accepting the hypothesis. If we now consider a sample statistic S and if the sampling of S is approximately normal, then we can be confident of finding the mean μ_S of the sampling distribution of S in the interval $S - 2\sigma_S$ to $S + 2\sigma_S$ 95.45% of the time or in the interval $S - 1.96\sigma_S$ to $S + 1.96\sigma_S$ 95% of the time. These intervals are called confidence intervals. The end numbers of these intervals $S \pm 1.96\sigma_S$ are called confidence limits.

We can now proceed considering the statistical significance of a correlation coefficient. A correlation coefficient r_{jk} computed from the measurements on variables Z_j and Z_k can be considered as an estimate of the true population correlation coefficient, denoted by ρ_{jk}. The measurements on Z_j and Z_k, taken as pairs (Z_{ji}, Z_{ki}), $i=1,\ldots,N$, are considered a sample from the population of all possible such pairs. Since two variables are involved, the population is called bivariate. We assume that it has a bivariate normal distribution. We are interested in whether the observed correlation coefficient differs significantly from an expected result. This obviously depends on the sample size N. The larger N will be the better will be the estimate of the true population coefficient by the sample coefficient. Then a statement about the error or precision of the estimate is called its reliability. In order to find out about statistical significance of r_{jk}, we have to test two hypotheses, namely that ρ_{jk} is zero or is not zero. To be able to test these hypotheses, we have to know the sampling distribution of r_{jk} for each case. For the hypothesis $\rho_{jk} = 0$, this distribution is symmetric and can be described by a statistic involving Student's t-distribution. If $\rho_{jk} \neq 0$, the sampling distribution of r_{jk} is skewed. Then Fisher's Z-transformation can be employed to transform the skewed distribution into one which is approximately normal. Let us express now these considerations mathematically:

1. Hypothesis $\rho_{jk} = 0$:

$$t = \frac{r_{jk}\sqrt{N-2}}{\sqrt{1-r_{jk}^2}}$$

has Student's t-distribution with N - 2 degrees of freedom. The hypothesis is accepted at a predetermined level of significance, if, for r_{jk}, t is computed to be less than the t-value read from Student's t-distribution table at the given level of significance and at given degrees of freedom.

2. Hypothesis $\rho_{jk} \neq 0$: The distribution of r_{jk} is transformed by Fisher's Z-transformation. We obtain:

$$Z = \frac{1}{2}\log_e\left(\frac{1+r_{jk}}{1-r_{jk}}\right)$$

with mean

$$\mu_Z = \frac{1}{2}\log_e\left(\frac{1+\rho_{jk}}{1-\rho_{jk}}\right)$$

and standard deviation

$$\sigma_Z = \frac{1}{\sqrt{N-3}} \quad .$$

Within the context we will be especially interested to find 95% confidence limits for r_{jk}. We proceed to do so by first testing the hypothesis that for a given correlation coefficient r_{jk} the true population coefficient ρ_{jk} is zero. If the hypothesis is rejected we are able to compute confidence limits.

We have learned that we can be 95% confident to find the mean μ_Z in the interval $Z \pm 1.96\sigma_Z$. It is

$$Z \pm 1.96\sigma_Z = \frac{1}{2}\log_e\left(\frac{1+r_{jk}}{1-r_{jk}}\right) \pm 1.96\frac{1}{\sqrt{N-3}}$$

which yields the confidence limits

$$\mu_Z = \frac{1}{2} \log_e \left(\frac{1 + r_{jk}}{1 - r_{jk}} \right) \pm 1.96 \frac{1}{\sqrt{N - 3}} \text{ for } \mu_Z .$$

Since

$$\mu_Z = \frac{1}{2} \log_e \left(\frac{1 + \rho_{jk}}{1 - \rho_{jk}} \right) ,$$

we are able to find confidence limits for ρ_{jk} from

$$\frac{1}{2} \log_e \left(\frac{1 + \rho_{jk}}{1 - \rho_{jk}} \right) = \frac{1}{2} \log_e \left(\frac{1 + r_{jk}}{1 - r_{jk}} \right) \pm 1.96 \frac{1}{\sqrt{N - 3}} .$$

$$\log_e \left(\frac{1 + \rho_{jk}}{1 - \rho_{jk}} \right) = \log_e \left(\frac{1 + r_{jk}}{1 - r_{jk}} \right) \pm 3.92 \frac{1}{\sqrt{N - 3}} .$$

Since

$$x = \log_e e^x ,$$

$$\log_e \left(\frac{1 + \rho_{jk}}{1 - \rho_{jk}} \right) = \log_e \left(\frac{1 + r_{jk}}{1 - r_{jk}} \right) \pm \log_e e^{3.92 \frac{1}{\sqrt{N - 3}}} .$$

Call

$$e^{3.92 \frac{1}{\sqrt{N - 3}}} = A(N),$$

then

$$\frac{1 + \rho_{jk}}{1 - \rho_{jk}} = \left(\frac{1 + r_{jk}}{1 - r_{jk}} \right) A(N) \qquad (5)$$

and

$$\frac{1 + \rho_{jk}}{1 - \rho_{jk}} = \left(\frac{1 + r_{jk}}{1 - r_{jk}}\right) \Big/ A(N) \tag{6}$$

from Equation 5 we obtain

$$\rho_{jk} = \frac{[A(N) - 1] + [A(N) + 1]r_{jk}}{[A(N) + 1] + [A(N) - 1]r_{jk}} \tag{5a}$$

from Equation 6 we obtain

$$\rho_{jk} = \frac{\left[\frac{1}{A(N)} - 1\right] + \left[\frac{1}{A(N)} + 1\right]r_{jk}}{\left[\frac{1}{A(N)} + 1\right] + \left[\frac{1}{A(N)} - 1\right]r_{jk}}$$

$$= \frac{[1 - A(N)] + [1 + A(N)]r_{jk}}{[1 + A(N)] + [1 - A(N)]r_{jk}} \tag{6a}$$

Equations 5a and 6a furnish the confidence limits for the correlation coefficient r_{jk}, whose corresponding population coefficient is ρ_{jk}.

One is referred to Spiegel (Reference 20) as a reference for this subsection.

2.6 PEARSON'S PRODUCT-MOMENT CORRELATION COEFFICIENT DERIVED FROM INCOMPLETE DATA

Let us assume that we have n variables and N individuals, on which observations are taken. It can quite often happen in practice that, for some reason, observations for a variable can be taken only for some of the N individuals. There are several ways to compute a Pearson product-moment correlation coefficient on the basis of a different number of observations for each variable. In the following, three methods are described. Each time the basic formula (in terms of raw scores):

$$r_{jk} = \frac{\sum_{i=1}^{N} X_{ji}X_{ki} - \frac{\sum_{i=1}^{N} X_{ji} \sum_{i=1}^{N} X_{ki}}{N}}{N\sqrt{\frac{\sum_{i=1}^{N} X_{ji}^2}{N} - \left[\frac{\sum_{i=1}^{N} X_{ji}}{N}\right]^2} \cdot \sqrt{\frac{\sum_{i=1}^{N} X_{ki}^2}{N} - \left[\frac{\sum_{i=1}^{N} X_{ki}}{N}\right]^2}}$$

is adjusted for the situation of incomplete data by particular means.

1. Method: SRL-Routine for the Computation of r_{jk}.

The correlation coefficient is computed on the basis of simultaneously existing data points for the two variables X_j and X_k.

$$r_{jk} = \frac{\sum_{i=i^*} X_{ji}X_{ki} - \frac{\sum_{i=i^*} X_{ji} \cdot \sum_{i=i^*} X_{ki}}{N^*}}{N^*\sqrt{\frac{\sum_{i=i^*} X_{ji}^2}{N^*} - \left[\frac{\sum_{i=i^*} X_{ji}}{N^*}\right]^2} \cdot \sqrt{\frac{\sum_{i=i^*} X_{ki}^2}{N^*} - \left[\frac{\sum_{i=i^*} X_{ki}}{N^*}\right]^2}},$$

where N^* is the number of data points, which exist for X_j and X_k simultaneously. The index i^* picks from the set $\{1,\ldots,N\}$ those numbers, which are accounted for in N^*.

Rewriting the formula above, we obtain

$$r_{jk} = \frac{\frac{\sum_{i=i^*} X_{ji}X_{ki}}{N^*} - \frac{\sum_{i=i^*} X_{ji}}{N^*} \cdot \frac{\sum_{i=i^*} X_{ki}}{N^*}}{\sqrt{\frac{\sum_{i=i^*} X_{ji}^2}{N^*} - \left[\frac{\sum_{i=i^*} X_{ji}}{N^*}\right]^2} \sqrt{\frac{\sum_{i=i^*} X_{ki}^2}{N^*} - \left[\frac{\sum_{i=i^*} X_{ki}}{N^*}\right]^2}}.$$

We form the means

$$\bar{X}_j^* = \frac{\sum_{i=i^*} X_{ji}}{N^*}$$

and

$$\bar{X}_k^* = \frac{\sum_{i=i^*} X_{ki}}{N^*}$$

For the computation of the means only those observations of X_j and X_k are taken into account, where observations exist for X_j and X_k simultaneously.

Then the correlation coefficient formula reduces to

$$r_{jk} = \frac{\frac{\sum_{i=i^*} X_{ji}X_{ki}}{N^*} - \bar{X}_j^* \bar{X}_k^*}{\sqrt{\frac{\sum_{i=i^*} X_{ji}^2}{N^*} - \bar{X}_j^{*2}} \sqrt{\frac{\sum_{i=i^*} X_{ki}^2}{N^*} - \bar{X}_k^{*2}}} \qquad (7)$$

A computer program for the above outlined computations can also be found in Reference 21.

2. Method: Computation of r_{jk} by Making Use of all Available Data.

A consequence of the first method is that, in computing only from simultaneously existing data points for both variables, valuable information is neglected, especially in computing the means. The means are based on smaller data sets than available and may therefore not as precisely describe the true population means as would be possible by use of all available data. Therefore the following method of computation of r_{jk}, which takes into account all available data for the computation of means and standard deviations, is suggested.

Let

$$\bar{X}_j = \sum_i \frac{X_{ji}}{N1} \quad ,$$

$$\bar{X}_k = \sum_i \frac{X_{ki}}{N2} \quad ,$$

where N1 accounts for the data points existing for X_j, N2 for the data points of X_k. In both cases summation is done over the set of existing data points of the variables under consideration.

$$r_{jk} = \sum_{i=i^*} \frac{Z_{ji} Z_{ki}}{N^*}$$

The correlation coefficient is a dot product of two standardized variables divided by the number of points taken into consideration. Since the product between two points has only meaning if neither of the points is missing, the summation will be done over the number of simultaneously existing points ($i=i^*$) and the sum will be divided by N^*, accounting for this number of points.

$$r_{jk} = \frac{\sum_{i=i^*} (X_{ji} - \bar{X}_j)(X_{ki} - \bar{X}_k)}{N^* \sigma_j \sigma_k}$$

$$r_{jk} = \frac{\sum_{i=i^*} (X_{ji} - \bar{X}_j)(X_{ki} - \bar{X}_k)}{N^* \sqrt{\sum_i \frac{(X_{ji} - \bar{X}_j)^2}{N1}} \sqrt{\sum_i \frac{(X_{ki} - \bar{X}_k)^2}{N2}}}$$

$$= \frac{\sum_{i=i^*} X_{ji}X_{ki} - \bar{X}_j \sum_{i=i^*} X_{ki} - \bar{X}_k \sum_{i=i^*} X_{ji} + N^* \bar{X}_j \bar{X}_k}{N^* \sqrt{\sum_i \frac{X_{ji}^2}{N1} - \bar{X}_j^2} \sqrt{\sum_i \frac{X_{ki}^2}{N2} - \bar{X}_k^2}} .$$

Let

$$\sum_{i=i^*} \frac{X_{ji}}{N^*} = \bar{X}_j^*$$

and

$$\sum_{i=i^*} \frac{X_{ki}}{N^*} = \bar{X}_k^* \quad ,$$

as in the case of the first method and obtain

$$r_{jk} = \frac{\sum_{i=i^*} \frac{X_{ji}X_{ki}}{N^*} - \bar{X}_j\bar{X}_k^* - \bar{X}_k\bar{X}_j^* + \bar{X}_j\bar{X}_k}{\sqrt{\sum_i \frac{X_{ji}^2}{N1} - \bar{X}_j^2}\ \sqrt{\sum_i \frac{X_{ki}^2}{N2} - \bar{X}_k^2}}$$

$$r_{jk} = \frac{\sum_{i=i^*} \frac{X_{ji}X_{ki}}{N^*} - \bar{X}_j\bar{X}_k^* - \bar{X}_k\bar{X}_j^* + \bar{X}_j\bar{X}_k + \bar{X}_j\bar{X}_k^* - \bar{X}_j\bar{X}_k^*}{\sqrt{\sum_i \frac{X_{ji}^2}{N1} - \bar{X}_j^2}\ \sqrt{\sum_i \frac{X_{ki}^2}{N2} - \bar{X}_k^2}}$$

$$r_{jk} = \frac{\sum_{i=i^*} \frac{X_{ji}X_{ki}}{N^*} - \bar{X}_j^*\bar{X}_k^* + (\bar{X}_j - \bar{X}_j^*)(\bar{X}_k - \bar{X}_k^*)}{\sqrt{\sum_i \frac{X_{ji}^2}{N1} - \bar{X}_j^2}\ \sqrt{\sum_i \frac{X_{ki}^2}{N2} - \bar{X}_k^2}} \quad . \qquad (8)$$

If $N1 = N2$, then $N^* = N1$ and $N^* = N2$, and $\bar{X}_k = \bar{X}_k^*$ and $\bar{X}_j = \bar{X}_j^*$. In this case Equations 7 and 8 are identical.

3. Method: Substitution for Missing Data Points

Another means one can think of as a solution to the problem of computing r_{jk} from incomplete data is the means of inserting some value for the missing data points of the variables. The values, which suggest themselves for substitution, are the statistical means of the variables. Since all sums are then taken over N variables the formula

$$r_{jk} = \frac{\sum_{i=1}^{N} X_{ji}X_{ki} - \frac{\sum_{i=1}^{N} X_{ji} \sum_{i=1}^{N} X_{ki}}{N}}{\sqrt{\sum_{i=1}^{N} \frac{X_{ji}^2}{N} - \left[\frac{\sum_{i=1}^{N} X_{ji}}{N}\right]^2} \cdot \sqrt{\frac{\sum_{i=1}^{N} X_{ki}^2}{N} - \left[\frac{\sum_{i=1}^{N} X_{ki}}{N}\right]^2}}$$

can be reduced to

$$r_{jk} = \frac{\sum \frac{X_{ji}X_{ki}}{N} - \bar{X}_j\bar{X}_k}{\sqrt{\sum_{i=1}^{N} \frac{X_{ji}^2}{N} - \bar{X}_j^2} \sqrt{\sum_{i=1}^{N} \frac{X_{ki}^2}{N} - \bar{X}_k^2}} \qquad (9)$$

where means

$$\bar{X}_j = \sum_i \frac{X_{ji}}{N1}, \quad \bar{X}_k = \sum_i \frac{X_{ki}}{N2}$$

(defined as in the 2. method) are substituted for missing data points in variables X_j, X_k respectively.

The advantage of this last method is, that in the final computation of the correlation coefficient only one sample size N is used. This can be of particular importance, if the correlation coefficient is later on used for statistical considerations, which are based on sample size N.

Remarks:

A. One general remark can be made concerning the three discussed methods: If the total number of observations is large, some missing data points will not affect the correlation coefficient, computed by the three methods, very much. This is based on the fact, that the mean, with large sample size N, gets nearly stable.

B. It is important to know, what to do when data points are missing. An example can be given reflecting this importance. A correlation matrix

was computed by using Equation 7. It happened that for the computation of one r_{ab} there were only a very few simultaneous observations on variables a and b, while for all other computations the number of simultanious observations was much larger and almost equal. This fact was not observed when the correlation matrix was established. At another step of computation later on, however, it was exhibited how influential the different numbers of observations were: The correlation matrix R was not Gramian (symmetric and all principal minors greater than or equal to zero) any more, what it should have been according to the way it was derived as $R = ZZ^T/N$. In using Equation 7--as well as Equation 8--the N is different, however, for each element of R. Only by using Equation 9 one computes all elements on the basis of the same sample size N. This is an advantage with respect to preserving Gramian properties. On the other hand, substitution of means for missing data points may disturb the true relations of the variables too much, so that a later factor analysis of the correlation matrix may not reflect the true intercorrelations among the variables any more. This suggests that the product--moment correlation coefficient should be computed by either Equation 7 or Equation 8.

Example: As an example for the three considered methods, 130 pairs of adult male height and weight measurements were selected. Using Pearson's product--moment correlation formula the correlation coefficient of the two variables, on the basis of 130 pairs of measurements, is computed to be 0.484.

To exhibit the three formulae for different degrees of missing data, three random samples were drawn from the sample of 130 measurements.

a. A random subset was drawn, such that 75% of all available data were used, 50% in complete pairs (height, weight), so that 50% of complete data pairs were missing.

b. Next, a random subset was drawn, such that 85% of all available data were used, 70% in complete pairs, so that 30% of complete data pairs were missing.

c. In the same manner, a random sample was drawn, such that 95% of all available data were used, 90% in complete pairs, so that 10% of complete data pairs were missing.

The results of the computation are listed below.

	Missing data pairs		
Computation by	50%	30%	10%
Equation 7	0.544	0.508	0.492
Equation 8	0.557	0.537	0.500
Equation 9	0.383	0.437	0.473

It is seen that the values computed by Equations 7 and 8 converge from above and the values computed by Equation 9 from below to the "true" value 0.484.

2.7 MULTIVARIATE CORRELATION--PARTIAL AND MULTIPLE CORRELATION COEFFICIENTS

Since in later sections we shall use the multiple correlation coefficient we will briefly consider it and also the partial correlation coefficient, for the sake of completeness, in this subsection.

To help clarify the nature of both coefficients let us consider the following problem. Assume that the variables (scores on them are given) stature, intelligence, and quickness of decision contribute to leadership. We term the factor leadership the dependent variable and the other three factors independent variables. Then, if we determine the correlation of the dependent variable with one of the independent variables, while the influence of the other independent variables is held constant, we determine what is named the coefficient of partial correlation between the two variables under consideration.

Mathematically we can express the above problem in the form of a regression equation. Let x_1, x_2, x_3, in deviate form, represent the independent variables and let $\bar{x}_0$ represent the dependent variable, which is estimated from the independent variables. The equation

$$\bar{x}_0 = b_{01.23}x_1 + b_{02.13}x_2 + b_{03.12}x_3$$

is called a regression equation of x_0 on x_1, x_2, x_3, the b's being constants. The graph x_0 versus x_3, for example, is a straight line with slope $b_{03.12}$. In this coefficient the indices left of the dot show the two related variables, while the indices right of the dot show the variables held constant. With respect to the fact that x_0 varies due partially to the variation in x_1 and due partially to the variation in x_2 and x_3 the b-coefficients are called partial regression coefficients.

Generalizing the above, a regression equation of x_0 on $x_1, x_2, \ldots, x_k$ can be written as

$$x_0 = b_{01.23\ldots k}x_1 + b_{02.134\ldots k}x_2 + \cdots + b_{0k.12\ldots(k-1)}x_k$$

The partial regression coefficients can, if necessary, be computed by the Doolittle Method. Then the partial correlation coefficients are easily developed from the notion of partial regression coefficients. When, in 2.2, we developed Pearson's product--moment correlation coefficient, we learned that the correlation coefficient r_{jk} is given by the slope corrected for the different measures of variability of x_j and x_k: $r_{jk} = a\sigma_k/\sigma_j$. Hence, a partial regression coefficient is the slope of the line relating the paired measures of a dependent variable and one independent variable, when the influence of the other independent variables has been excluded from consideration but when the units of measurement are not necessarily of equal variability. Corresponding to the development of Pearson's product--moment correlation coefficient let us now develop the partial correlation coefficient. Let $b_{0i.123\ldots)i(\ldots k}$ denote the partial regression coefficient between the dependent variable x_0 and the independent variable x_i, while the independent variables $x_1, x_2, \ldots, x_k$, excluding x_i, are held constant. Let $\sigma_{0,123\ldots k}$ and $\sigma_{i.12\ldots)i(\ldots k}$ stand for the standard deviation of variables x_0 and x_i when the effects of variables $x_1, x_2, \ldots, x_k$, and $x_1, x_2, \ldots)x_i(\ldots, x_k$ have been ruled out. Let $r_{0i.12\ldots)i(\ldots k}$ denote the partial correlation coefficient. Then

$$r_{0i.12...)i(...k} = b_{0i.12...)i(...k} \cdot \frac{\sigma_{i. \quad ...)i(...k}}{\sigma_{0.12...k}}$$

and

$$r_{i0.12...)i(...k} = b_{i0.12...)i(...k} \quad \frac{\sigma_{0.12...k}}{\sigma_{i.12...)i(...k}}$$

Since $r_{0i.12...)i(...k} = r_{i0.12...)i(...k}$ we define the partial correlation coefficient as:

$$r_{0i.12...)i(...k}$$

$$= \sqrt{b_{0i.12...)i(...k} \frac{\sigma_{i.12...)i(...k}}{\sigma_{0.12...k}} b_{i0.12...)i(...k} \frac{\sigma_{0.12...k}}{\sigma_{i.12...)i(...k}}}$$

$$= \sqrt{b_{0i.12...)i(...k} \cdot b_{i0.12...)i(...k}}$$

The sign is the sign computed for $b_{0i.12...)i(...k}$ or $b_{i0.12...)i(...k}$, both being the same.

Referring again to our example in the beginning of this subsection, another question, we might be interested in, could be: What is the correlation between leadership and the three independent variables taken jointly? The coefficient which describes the relationship between the dependent variable and the independent variables, taken together, is called the multiple correlation coefficient.

For a certain individual we actually get a score on the dependent variable, call it Z_{0i} $(i=1,...,N)$ in standard form. On the other hand by the regression equation we estimate such a score. So the multiple correlation coefficient, denoted by $R_{0.12...k}$, is defined as the correlation between the observed Z_0 and the computed $\overline{Z}_0$,

$$R_{0.12...k} = \sum_{i=1}^{N} \frac{Z_{0i}\overline{Z}_{0i}}{N\sigma_{Z_0}\sigma_{\overline{Z}_0}}$$

with $\overline{Z}_0 = \beta_{01.23...k}Z_1 + \cdots + \beta_{0k.12...k-1}Z_k$. Some computation done on the above equations yields

$$R_{0.12...k} = \sqrt{\beta_{01.23...k}r_{01} + \cdots + \beta_{0k.123...(k-1)}r_{0k}}; \; r_{01},...,r_{0k}$$

using Pearson's product--moment correlation coefficients.

Section III

THE CORRELATION MATRIX

3.1 INTRODUCTION

Since almost every factor analytic technique begins with a correlation matrix, properties of correlation matrices and techniques concerned only with correlation matrices are presented in this section.

Those theorems and definitions from eigenvalue theory which are particularly applicable to correlation matrices and which will be needed in factor analyses are presented in subsection 3.2. Subsection 3.3 contains a definition of a correlation matrix along with those properties which are important to factor analysis. Subsection 3.4 concludes the section with a presentation of scaling techniques based on sample size.

3.2 EIGENVALUE-EIGENVECTOR THEORY

In this subsection we will consider those definitions and theorems from the eigenvalue-problem theory which are necessary for and used in the development of factor analysis.

Let us first state the eigenvalue problem.

Consider the following algebraic problem: Given a matrix A of order n. Determine a scalar λ and an n-dimensional nonzero vector x, such that

$$Ax = \lambda x$$

(A, x can be over a complex field, λ a complex number).

Definition 3.1: The above problem is defined as the eigenvalue problem. The eigenvalue problem can be rewritten as

$$(A - \lambda I_n)x = 0.$$

This system of n homogeneous linear equations in n unknowns has nontrivial solutions, if and only if the determinant of the matrix of coefficients vanishes, i.e.,

$$d(A - \lambda I_n) = \begin{vmatrix} a_{11}-\lambda & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn}-\lambda \end{vmatrix} = 0.$$

Expanding the determinant we obtain a polynomial in λ of degree n, denoted by $\phi(\lambda)$, so that the requirement is $\phi(\lambda) = 0$.

Definition 3.2: The equation $\phi(\lambda) = 0$ is called the characteristic equation. The n roots of $\phi(\lambda)$ are named the eigenvalues of the matrix A. Associated with each such eigenvalue λ_i is a vector x_i, named an eigenvector of A.

Completing the statement of the problem we have the following theorem:

Theorem 3.1: The equation $Ax = \lambda x$ has nontrivial solutions x iff λ is an eigenvalue of A.

Example

Let A be:

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

The eigenvalue problem is

$$\left[\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}\right]\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0.$$

The characteristic equation is

$$\phi(\lambda) = \det \begin{pmatrix} 1-\lambda & 2 \\ 2 & 1-\lambda \end{pmatrix} = (1 - \lambda)^2 - 4 = 0,$$

with solutions $\lambda_1 = 3$ and $\lambda_2 = -1$, the eigenvalues of the problem. With each of the two eigenvalues is associated an eigenvector. For $\lambda_1 = 3$ and $\lambda_2 = -1$ the system $(A - \lambda I_n)x = 0$ each time reduces to a single equation:

$\lambda_1 = 3$

$$\begin{pmatrix} 1-3 & 2 \\ 2 & 1-3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -2 & 2 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

gives $x_1 - x_2 = 0$

$\lambda_2 = -1$

$$\begin{pmatrix} 1+1 & 2 \\ 2 & 1+1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_1 \end{pmatrix}$$

gives $x_1 + x_2 = 0.$

The complete solution set is then described by

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = k_1 \begin{pmatrix} 1 \\ +1 \end{pmatrix} \qquad \text{for} \quad x_1 - x_2 = 0,$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = k_2 \begin{pmatrix} 1 \\ -1 \end{pmatrix} \qquad \text{for} \quad x_1 + x_2 = 0.$$

Let us further restrict ourselves to real symmetric matrics A, since the matrics we deal with in factor analysis (the correlation matrices) are of this kind.

Let us consider the characteristic equation $\phi(\lambda)$, stating the following:

Theorem 3.2: The coefficient of $\lambda^r (r \leq n)$ in $\phi(\lambda)$ is $(-1)^r$ times the sum of the principal minors of order $n-r$ of A. In particular, the coefficient of λ^n is $(-1)^n$, the constant term is det A.

In the special case, where $r = n-1$, the coefficient of λ^{n-1} is

$$(-1)^{n-1}(a_{11} + a_{22} + \quad + a_{nn}) = (-1)^{n-1} \sum_{i=1}^{n} a_{ii}$$

where

$$\sum_{i=1}^{n} a_{ii}$$

is called the trace of A.

It will be useful to know the following two theorems about the roots of the characteristic equation:

Theorem 3.3: If $\bar{\lambda}$ is a simple root of $\phi(\lambda) = 0$, then the rank of $(A - \bar{\lambda}I)$ is $n-1$.

Theorem 3.4: If $\bar{\lambda}$ is an r-fold root of $\phi(\lambda)=0$, then the rank of $(A - \bar{\lambda}I)$ is $n-r$. [A root $\bar{\lambda}$ is called r-fold, if $(\lambda-\bar{\lambda})$ is contained in $\phi(\lambda)$ r times. A root which is not an r-fold root is called a simple root.]

Let us next consider some results about the eigenvalues and eigenvectors.

Theorem 3.5: The eigenvalues of a real symmetric matrix are all real.

Theorem 3.6: Eigenvectors associated with the eigenvalues of a real symmetric matrix have all real components.

Theorem 3.7: Eigenvectors associated with distinct eigenvalues of a real symmetric matrix A are orthogonal.

Let us now put one more restriction on the matrix A, namely the restriction that all its elements shall be greater than zero.

Theorem 3.8: Let all elements of the real symmetric matrix A be positive. Then A has always an eigenvalue λ, which is real and positive, which is a simple root of the characteristic equation and which is not exceeded in modulus by any other eigenvalue. The eigenvector corresponding to λ has positive components and is essentially unique (up to scale factors).

(The theorem is due to Perron (for the proof see Reference 22). It can be extended to so-called irreducible matrices, which case will be considered, though, in the context.)

3.3 DEFINITIONS AND PROPERTIES

We begin this subsection with

Definition 3.3: A correlation matrix R is a square matrix where each element r_{ij} is the correlation between the variables Z_j and Z_k.

In the sequel, it is assumed that Pearson's product-moment correlation is used.

The most important properties of a correlation matrix from the point of view of factor analysis are included in the statement that a correlation matrix is Gramian. A Gramian matrix may be defined by

Definition 3.4: Let R be symmetric. Then R is called Gramian, if it satisfies any one of the following equivalent conditions:

1. R is positive semi-definite.
2. R has all non-negative eigenvalues.
3. R can be represented by the matrix product AA^T.
4. R has non-negative principal minors.
5. The inner product, $(RX|X) \geq 0$, for all X.

Obviously the correlation matrix R is Gramian since it is obtained by the product of score matrics,

$$R = \left(\frac{Z}{\sqrt{N}}\right) \quad \left(\frac{Z}{\sqrt{N}}\right)^T,$$

Where N is the number of observations.

3.4 SCALING TECHNIQUES BASED ON SAMPLE SIZE

Two correlation matrices with identical elements will, of course, yield identical factor analyses. If it were the case that identical elements had different significance levels, these differences in reliability would not appear in the factors. Thus in order for a factor analysis to reflect the significance of the correlation coefficients,

the correlation coefficients should be scaled in accordance with their significance.

As used in Section 2.5, the variable

$$Z_r = \frac{1}{2} \ln \frac{1 + r}{1 - r}, \text{ Fisher's Z-transformation,} \qquad (1)$$

is approximately normally distributed with

$$s = \frac{1}{\sqrt{N - 3}} \qquad (2)$$

an unbiased estimate of the standard deviation where N is the size of the random sample used in computing the correlation coefficient, r. Then, for 95% of the samples, the variable

$$Z_\rho = Z_r - \frac{1.96}{\sqrt{N - 3}} \qquad (3)$$

will be less than the true population variable. Thus, Equation 3 may be used to obtain a scaled correlation coefficient ρ given the observed correlation coefficient r. ρ will have on the average one chance out of twenty of exceeding the true population correlation coefficient. The probability may be adjusted by changing the numerator of the second term in Equation 3.

Equation 1 may be used to solve for ρ as a function of r by substitution into Equation 3. We obtain (as derived in Section 2.5):

$$\rho = \frac{[1-A(N)] + [1+A(N)]r}{[1+A(N)] + [1-A(N)]r} \qquad (4)$$

Equation 4 is the formula to be used to scale observed correlation coefficients, r.

Since each element of a correlation matrix $R = (r_{ij})$ is a correlation coefficient, Equation 4 may be written as

$$\rho_{ij} = \frac{[1-A(N)] + [1+A(N)]r_{ij}}{[1+A(N)] + [1-A(N)]r_{ij}}$$

for application to correlation matrices with observed correlations, r_{ij}.

Figure 1 shows curves relating r and ρ for various sample sizes. A chart to be used in the same manner as the chart presented here appears in Ezekial and Fox, Methods of Correlation and Regression Analysis (Reference 23, p. 294). However the shape of the curves in the chart differs from those presented here, and the derivation of the chart is not given.

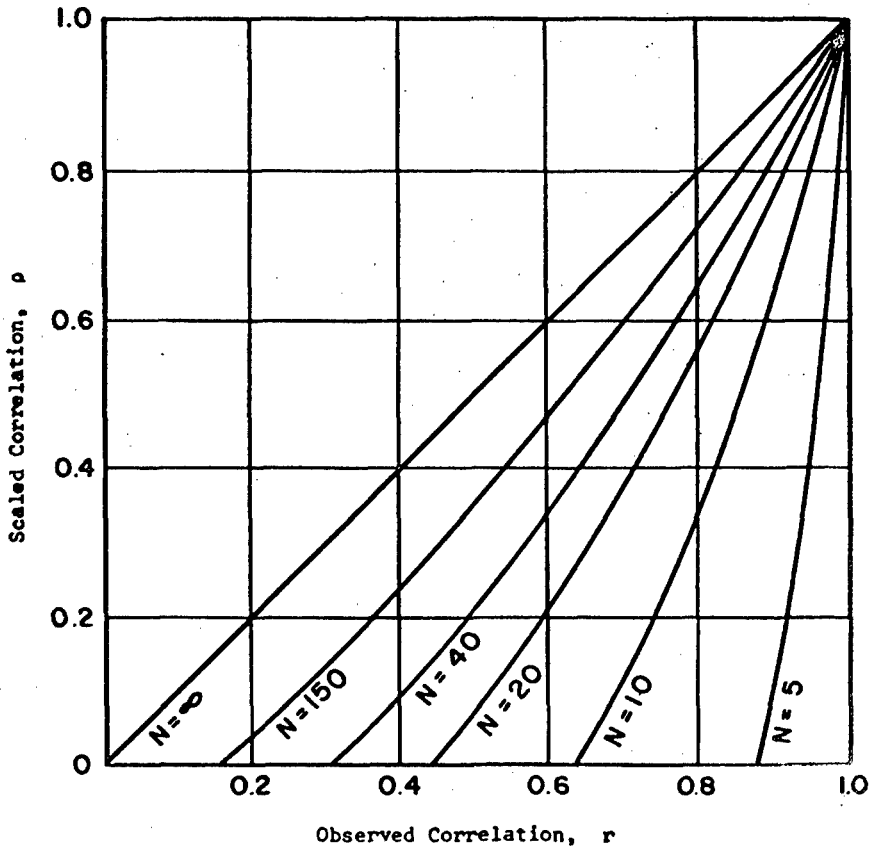


Figure 1. Correlation Coefficient Scaling Chart

Section IV

TECHNIQUES OF FACTOR ANALYSIS

4.1 INTRODUCTION

After the correlation matrix, which furnishes the basic material for a factor analysis, has been investigated, one can proceed to consider techniques of factor analysis. This is done by briefly reviewing the model in 4.2 and by then discussing, in 4.3, the properties of the two most important and popular factor analysis methods, the centroid and the principal-factors methods. Starting with section 4.4 specific problems of factoring a correlation matrix are discussed; 4.4 presents a new technique to estimate communalities; 4.5 compares most of the important completeness criteria; 4.6 called "Eigenvalues and Their Bounds" suggests a way to an answer on the important question of the right sample size. The section ends with a brief discussion of factor scores in 4.7.

4.2 REVIEW OF THE MODEL

In this subsection the model will be presented in greater detail stating basic definitions and equations.

We begin with the two basic theorems of factor analysis:

Theorem 4.1: For every correlation matrix R there exists a corresponding factor matrix F such that

$$FF^T = R.$$

Furthermore,

Theorem 4.2: There exists an infinite number of factor matrices F which reproduce any given correlation matrix R.

The problem, then, is not only to find an F, but to find the F that satisfies a given set of initial conditions which are, more often than not, subjective decisions and boundary criteria. The solution of the factor analysis problem consists of two basic steps:

1. Factoring problem - factor a given R into a factor matrix with an arbitrary reference frame.

2. Rotational problem - rotate the arbitrary reference frame into a "preferred" or "simplifying" position.

Factorial methods were developed primarily for the purpose of investigating and identifying the principal dimension or categories of mentality and thus are plagued by the non-mathematical justifications which are used to evaluate them. A technique infallible to a psychologist can be worthless to the engineer grading castings or a company rating its employees. Consequently, some of the basic definitions and techniques are given next using mathematical notation while comments on reliability and practicality for application result from longhand factor interpretation. From References 2 and 24 come necessary basic definitions and equations. It is the purpose of factor analysis to represent a variable X_j in terms of several underlying factors, or, as Harman (Reference 2) states, "hypothetical constructs". There are various kinds of factors:

Common factors - involved in more than one variable

a. General factor - present in all variables

b. Group factor - present in more than one but not in all variables

Unique factors - involved in a single variable.

We now use the notation $F_1, F_2, \ldots, F_m$ for, say, m common factors and $U_1, U_2, \ldots, U_n$ for, say, n unique factors to express linearly any variable in terms of the factors as follows:

$$X_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + a_jU_j.$$

For a particular individual or observation we have

$$X_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} + \cdots + a_{jm}F_{mi} + a_jU_{ji}, \quad i = 1, \ldots, N.$$

The coefficients a_{jp} ($j = 1, \ldots, n$, $p=1, \ldots, m$) are the elements of the factor matrix and are referred to as the factor loadings composing the factor matrix

$$F = [a_{jp}]$$

The total variance of X_j can be divided into two parts, namely, that part which it shares with other variables and that part which is unique. For example.

$$\sigma_j^2 = \frac{1}{N}\left(a_{j1}{}^2 \sum F_{1i}^2 + \cdots + a_{jm}^2 \sum F_{mi}^2 + a_j^2 \sum U_{ji}^2\right.$$

$$\left. + 2a_{j1}a_{j2} \sum F_{1i}F_{2i} + \cdots + 2a_{jlm}a_j \sum F_{mi}U_{ji}\right)$$

where all summation limits are $i = 1, 2, \ldots, N$

If the variables are in standard form and the factors are uncorrelated,

$$1 = \sigma_j^2 = a_{j1}^2 + \cdots + a_{jm}^2 + a_j^2$$

The terms on the right represent portions of the variance ascribable to the factors (i.e., a_{j1}^2 is the contribution of F_1 to the unit variance of Z_j). The <u>total contribution</u> of a factor F_p to the variances of all variables is defined to be

$$V_p = \sum_{j=1}^{n} a_{jp}^2$$

Uniqueness can be further broken down into <u>specific</u>, S_j, and <u>error</u>, E_j, factors. Since error and specific factors are uncorrelated,

$$a_j^2 = b_j^2 + c_j^2$$

where b_j and c_j are the respective factor loadings of S_j and E_j. Therefore the total variance can be expressed

$$1 = h_j^2 + a_j^2 = h_j^2 + b_j^2 + c_j^2.$$

<u>Communalities</u>, Then, are defined as the common-factor variances of the variables.

A set of equations giving any set of varibles $\{Z_j\}$ in terms of the m common factors and one unique factor is sometimes called a factor pattern. Such a pattern can be presented in tabular form, e.g.

$$
\begin{array}{cccccc}
a_{11} & a_{12} & \cdots & a_{1m} & & a_1 \\
a_{21} & a_{22} & \cdots & a_{2m} & & a_2 \\
\vdots & & & & & \\
a_{n1} & a_{n2} & & a_{nm} & & a_n
\end{array}
$$

The number of common factors included in such a description of a variable is the variable complexity. A factor matrix which represents the total unit variance of each variable is the complete factor matrix. A factor matrix which represents only the common factor variance of each variable is the reduced factor matrix. A correlation matrix with ones in the diagonal elements is referred to as the complete correlation matrix.

A row of the factor matrix in relation to an origin and reference frame in (m + N) - space (factor space) will be called a variable vector. The re-orientation of this vector within the space constitutes the rotation problem. Techniques for factor rotation are discussed in Section V.

4.3 TYPES OF FACTOR SOLUTIONS

Ideally a factor solution displaying a minimum complexity (i.e., a common factor space of one dimension or two dimensions) is the goal of the factor analyst. Such a factor pattern might look like

a_{11} a_1

a_{21} a_2

a_{n_1} a_n

The uni-factor and two-factor solutions are examples of such theoretical entities. However, the factor analyst rarely will see data which can be accounted for so simply. Data which are not well behaved, requiring a complex network of correlated and uncorrelated common factors as well as a set of inconsistent unique factors, is the rule rather than the exception. Consequently, the factor analyst generally must first decide what he is looking for and then choose a technique which best suits his needs. A rather short list of factoring methods is at his disposal. As a matter of fact there are but two popular methods practiced, differentiated significantly by the number of calculations involved. Thus the centroid method for years has set the standard in hand computation techniques while the principal axes method has proved itself workable using high speed digital computers. Both methods can lead to multiple-factor solutions. A short discussion on each of these methods follows.

Centroid Method - The centroid method of factoring exhibits what Thurstone calls a "computational compromise" since the resulting factor loadings are not unique for a given R. Let us assume, then, that the original score matrix S consists of n vectors contained in m-space where m is the number of common factors. As is well known, the correlations between any two of the n variables are just the scalar products between them. To obtain a vector whose m components give the centroid of the points describing the set of common factors, we simply average the elements in the factor matrix approximately, or

$$\left(\frac{1}{n}\sum a_k \, , \frac{1}{n}\sum a_k \, , \ldots , \frac{1}{n}\sum a_{km}\right)$$

where $k = 1, 2, \ldots n$. We require the frame of reference to have an axis passing through the centroid thus reducing the centroid vector to

$$\left(\frac{1}{n}\sum a_k \, , 0 \, , 0 \, , \cdots , 0\right)$$

With a minimum of transposition utilizing the new axis, a general formula can easily be derived which gives the elements of the first factor loadings:

$$a_{j1} = \frac{S_j}{\sqrt{T}} \, , \quad j = 1, 2, \ldots, n$$

$$S_j = \sum_{k=1}^{n} r_{jk} \text{ and } T = \sum_{j=1}^{n} \sum_{k=1}^{n} r_{jk}$$

The residual matrix is then calculated

$$[r_{jk}] = [r_{jk} - a_{j1}a_{k1}]$$

and the next factor loadings are calculated using

$$a_{j2} = \frac{\epsilon_j S_{j1}}{\sqrt{T_1}} \, , \quad j = 1,2,\ldots,n \, ,$$

where $\epsilon_n = \pm 1$, depending on necessary matrix reflections.

Removal of the remaining factors follows the same pattern until the process is ended. Interestingly enough, no dependable technique exists to stop this sequence. However, this problem will be considered in subsection 4.5

Principal Axes Method - The principal axes method of factoring derives an ellipsoid representation where the axes of the ellipsoids correspond to the factors. The selection of the factors occurs such that their respective contributions to the communality decreases. In other words, the contribution of factor one to the total communality is maximum.

Therefore, $V = a_{11}^2 + a_{21}^2 + \ldots + a_{n1}^2$ is chosen as maximum under the conditions

$$r_{jk} = \sum_{p=1}^{m} a_{jp} a_{kp} \quad , \quad j,k = 1,2,\ldots, n.$$

Applying differential calculus to these conditions, the characteristic equation of the correlation matrix R is derived

$$\left| R - \lambda I \right| = 0.$$

Solutions of the characteristic equation are, of course, the eigenvalues of the matrix which have the following well known property, generally expressed:

$$\lambda_p = \sum_{j=1}^{n} a_{jp}^2$$

The set of eigenvectors $\{\alpha_{jp}\}$ corresponding to λ_p then are used to obtain the factor loadings of factor F_p:

$$a_{jp} = \frac{\alpha_{jp}\sqrt{\lambda_p}}{\sqrt{\alpha_{1p}^2 + \alpha_{2p}^2 + \cdots + \alpha_{np}^2}} \quad , \quad j=1,\ldots, n.$$

Of course, communalities must be estimated in this process and can affect the solution for a small number of variables. Since a decreasing amount of communality is extracted with each factor, an ϵ can be chosen such that $|\bar{h}^2 - h^2| \leq \epsilon$ completes the factoring where $\bar{h}^2$ is the derived approximated communality.

4.4 COMMUNALITY

In most standard factor analysis packages available in computation centers all over the country, there exist at least three alternatives in the estimation of communalities: insertion of ones in the diagonal of R; arbitary choice of h_j^2 as the largest r_{ij} in either the i^{th} row or the j^{th} column (this is the technique employed by Thurstone (Reference 25) in deriving human factors); or by using squared multiple correlations (see Section 2.7) as communalities. Harman (Reference 2, p.86) states the problem as follows:

> "Literally dozens of methods for estimating communalities have been proposed but none of them has been shown to be superior to any of the others on the basis of closer approximation to the "true" values. As a matter of fact none of the methods has been demonstrated to lead to minimal rank of the correlation matrix. The choice among the various methods of approximation is generally made on the basis of available computational facilities and the disposition of the investigator to employ that method which intuitively seems best to approach the concept of communality."

In this subsection a new technique to obtain communalities is presented. Let us first introduce in more detail the basic requirements for estimation of communalities.

For uncorrelated factors the communality, h_j^2, of the j^{th} variable is given by the sum of the squares of the common factor coefficients, viz.,

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2.$$

The elaboration of this statement has yielded further defining characteristics:

1. The communality may be defined as the squared multiple correlation of the given observed variable on the common factors.

2. The squared multiple correlation of the given variable on the remaining variables must be the lower bound to the communality (References 26 and 27).

3. The communality is the upper limit of this squared multiple correlation as the number of variables approaches infinity (Reference 27).

4. Since the communality is a variance, its upper limit is one.

5. Unique communalities can be obtained only when the rank of the matrix satisfies the following condition (References 28 and 25).

$$m \leq \frac{2n + 1 - \sqrt{8n + 1}}{2} .$$

These requirements yield properties which have been stated as follows (Reference 29):

1. The obtained communalities must be within the following boundaries:

$$0 \leq R_j^2 \leq h_j^2 \leq 1 .$$

2. The factor loading matrix should reproduce the reduced correlation matrix exactly.

3. Minimum rank should be attained.

4. The reduced correlation matrix should be Gramian.

When the principal factor method is used, properties 2 and 4 can be shown to be equivalent.

Guttman (Reference 30) has shown that diagonal values which reduce rank may not satisfy other requirements for communalities. The Heywood case (Reference 31) is the classic example. Moreover, the statement often made that the rank of any symmetric matrix with even random elements can always be reduced to a certain degree by choosing diagonal values has been shown to be false (Reference 29). The proof is based on the impossibility of assuring real solutions to systems of nonlinear equations with real coefficients. From intuitive considerations of experiment design, it is to be expected that the number of factors causing variance among the variables is even greater than the number of variables. Other minor factors cause variance in the measure of variables intended to measure major factors.

Thus the attempt to find diagonal values which reduce rank must end in only some sort of approximation. But rank reduction is basic to a parsimonious explanation of the variance of the variables, and different

approaches may be taken to the problem of approximating rank reduction. For example, one of the few communality estimates based on rationale, the method of triads, uses average diagonal values which seek to force determinants of submatrices approximately to zero. On the other hand, the refactoring method simply postulates the number of factors. Many so-called "estimates" of communality do not even consider rank reduction.

Then from the foregoing statements we may distill a refined definition of the communality problem:

> Find diagonal values h_j such that $0 \leq R_j^2 \leq h_j^2 \leq 1$, and such that the correlation matrix with these diagonal values is Gramian. Moreover with these diagonal values, a higher percentage of common factor variance is explained with fewer factors than with any other diagonal values.

A method for computing diagonal values which attempts to satisfy this definition is described in the sequel.

A. The Method

If a symmetric matrix A is bordered by the column U, the row U^*, and the scalar α, then the eigenvalues λ of

$$\tilde{A} = \begin{pmatrix} A & U \\ U^* & \alpha \end{pmatrix} \tag{1}$$

satisfy the equation,

$$\lambda - \alpha = \sum_{i=1}^{n} \frac{(U \mid X_i)^2}{\lambda - \lambda_i}, \tag{2}$$

where X_i is the unit eigenvector corresponding to the eigenvalue λ_i of the $n \times n$ matrix A (Reference 32, P. 27).

Since the rank of a matrix is the order of the matrix is the order of the matrix minus the number of zero eigenvalues, to reduce the rank we must have zero eigenvalues. Then, in view of Equation 2, a necessary condition for zero eigenvalues is that

$$\alpha = \sum_{i=1}^{n} \frac{(U \mid X_i)^2}{\lambda_i} . \tag{3}$$

When $\lambda_i = 0$, it can be shown that $(U \mid X_i) = 0$, thus

$$\lim_{\lambda_i \to 0} \frac{(U \mid X_i)^2}{\lambda_i} = 0 \quad .$$

Thus the terms in Equation 3 corresponding to zero eigenvalues of A may be elided.

The foregoing scheme will be used to find each diagonal element. Before presenting the algorithm formally, a theorem on transforming eigenvalues is needed. Two diagonal elements of R, R_{kk} amd R_{nn}, may be interchanged by the transformation

$$I_k R I_k$$

where I_k is the identity matrix with the k^{th} and n^{th} rows (or columns) interchanged. The use to be made of this transformation rests on the

Theorem 4.3: R and $I_k R I_k$ have the same eigenvalues.

Proof: R may be diagonalized by an orthogonal transformation P by

$$R = P^* \Lambda P$$

where Λ is the diagonal matrix of eigenvalues of R.
Now we may consider $R_k = I_k R I_k$. Using the facts that $I_k I_k = I$ and $I_k^* = I_k$ we have

$$\begin{aligned} R_k = I_k R I_k &= I_k P^* \Lambda P I_k \\ &= (P I_k)^* \Lambda (P I_k) \\ &= P_1^* \Lambda P_1 \end{aligned}$$

Thus R_k is diagonalized by the orthogonal matrix P_1 with the same diagonal eigenvalue matrix Λ. Moreover, the eigenvectors are also permuted since

$$RX = \lambda X \quad ,$$

$$I_k A (I_k I_k) X = \lambda I_k X \quad ,$$

$$(I_k R I_k)(I_k X) = \lambda(I_k X) ,$$

and

$$R_k(I_k X) = \lambda(I_k X) .$$

Then we may consider approaching the communality problem by choosing R_{kk} as follows:

$$R_{kk} = \sum_{i=1}^{n} \frac{(U_k \mid X_i)^2}{\lambda_i}, \quad k = 1, 2, \ldots, n$$

where

$$R_k' X_i = \lambda_i X_i ,$$

$$R_k = \begin{pmatrix} R_k' & U_k \\ U_k^* & R_{kk} \end{pmatrix} = I_k R I_k ,$$

and

$$(X_i \mid X_i) = 1 .$$

Both unities and squared multiple correlations have been used as initial diagonal values. After the diagonal values have been found for all k, the process is repeated until all diagonal values are stable. Diagonal values are used (i.e., replace old values) as soon as they are calculated. In practice, convergence is enhanced by omitting terms in Equation 3 for which $\lambda_i < \varepsilon = .05$. The final result of the method is a clustering of eigenvalues about zero. Thus there are small negative eigenvalues. For the sake of interpretation Gramian properties are not necessary. However when data reduction is the object of the factor analysis, Gramian properties may be restored by adding the absolute value of the negative eigenvalue with the largest absolute value to each element on the diagonal. In proof we write

$$R = P^* \Lambda P$$

where Λ is the diagonal matrix of eigenvalues of R.

Let Λ_n be the scalar matrix $\lambda_n I$;

then $$R - \Lambda_n = P^*\Lambda P - \Lambda_n .$$

But since $P^*P = I$ and commutes with Λ_n,

$$\begin{aligned} R - \Lambda_n &= P^*\Lambda P - P^*\Lambda_n P \\ &= P^*(\Lambda - \Lambda_n)P . \end{aligned}$$

Thus the elements of the diagonal matrix $\Lambda - \Lambda_n$ are the eigenvalues of $R - \Lambda_n$. But the elements of $\Lambda - \Lambda_n$ are all positive or zero. Therefore $R - \Lambda_n$ is Gramian. However, this method for forcing Gramian properties may lead to communalities larger than 1.

To better understand how the above bordering scheme may drive the eigenvalues of R to zero and thus reduce the rank, let us plot on the same graph each side of Equation 2 as a function of λ (see Figure 2). The solid graph is the right hand side, and the dotted graph is the left hand side of Equation 2. The eigenvalues λ of the bordered matrix occur at the intersections of the sets of curves. Notice that the eigenvalue of the smaller matrix always lies between two eigenvalues of the larger matrix. Each of the dotted lines corresponds to a different choice of α. The uppermost dotted line corresponds to the α chosen according to Equation 3, in which case we have an eigenvalue of zero. The observed effect of reapplying the algorithm after transformation of the matrix (i.e., finding a new diagonal element) is to shift positive eigenvalues to the left (closer to zero) and negative eigenvalues to the right (closer to zero). A formal deductive proof of convergence has not yet been found; however the success of the algorithm in solving the communality problem is exhibited in the following examples.

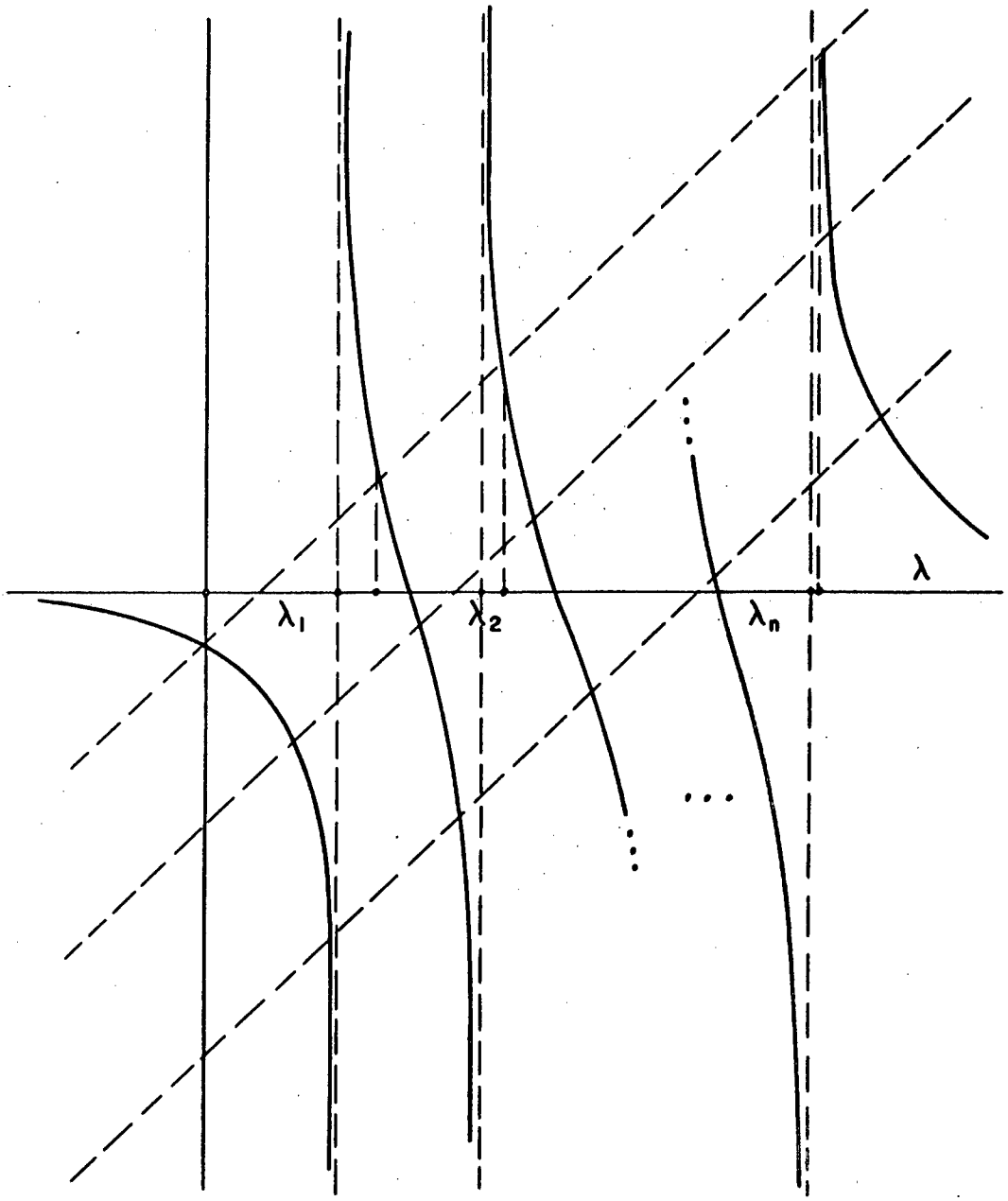


Figure 2. Plot of Equation 2

B. Examples

Example 1. Six Hypothetical Variables: Harman (Reference 2, p. 91) used these variables to illustrate communality estimates by various methods. The results of applying the method described in this paper are shown by plotting the eigenvalues of the 6 x 6 correlation matrix with calculated diagonal values (Figure 3). The eigenvalues obtained using unities and squared multiple correlation are also plotted in Figure 3. In this example and in every other application of the method described here, the following inequalities have held:

$$\lambda_i(R^2) < \lambda_i(d^2) < \lambda_i(1) \quad i = 1,\ldots, n$$

$$R_i^2 < d_i^2 < 1 \quad i = 1,\ldots, n$$

where d_i^2 are the calculated diagonal elements and R_i^2 are squared multiple correlations. The calculated diagonal elements are "true" communalities in the sense that the correlation matrix was constructed to attain rank two with these values. No communality estimate presented by Harman found these values.

Example 2. Thirteen Psychological Variables: The data for this example was also taken from Harman (Reference 2, p. 137). However these variables are experimental rather than hypothetical. The plot of the three sets of eigenvalues are shown in Figure 4. These variables were well chosen to illustrate three major factors as clearly seen in Figure 4. However, it would only be accidentally possible to find diagonal values which would yield ten zero eigenvalues (i.e., a rank three correlation matrix).

Example 3. 16 Hypothetical Variables: A 16 x 16 matrix was constructed by squaring a 16 x 4 matrix of random elements with normalized columns. Thus the 16 x 16 matrix was of rank 4 when the constructed diagonal elements were retained. The proposed method found these "true communalities" given the constructed matrix with unities on the diagonal. The plots of eigenvalues are shown in Figure 5.

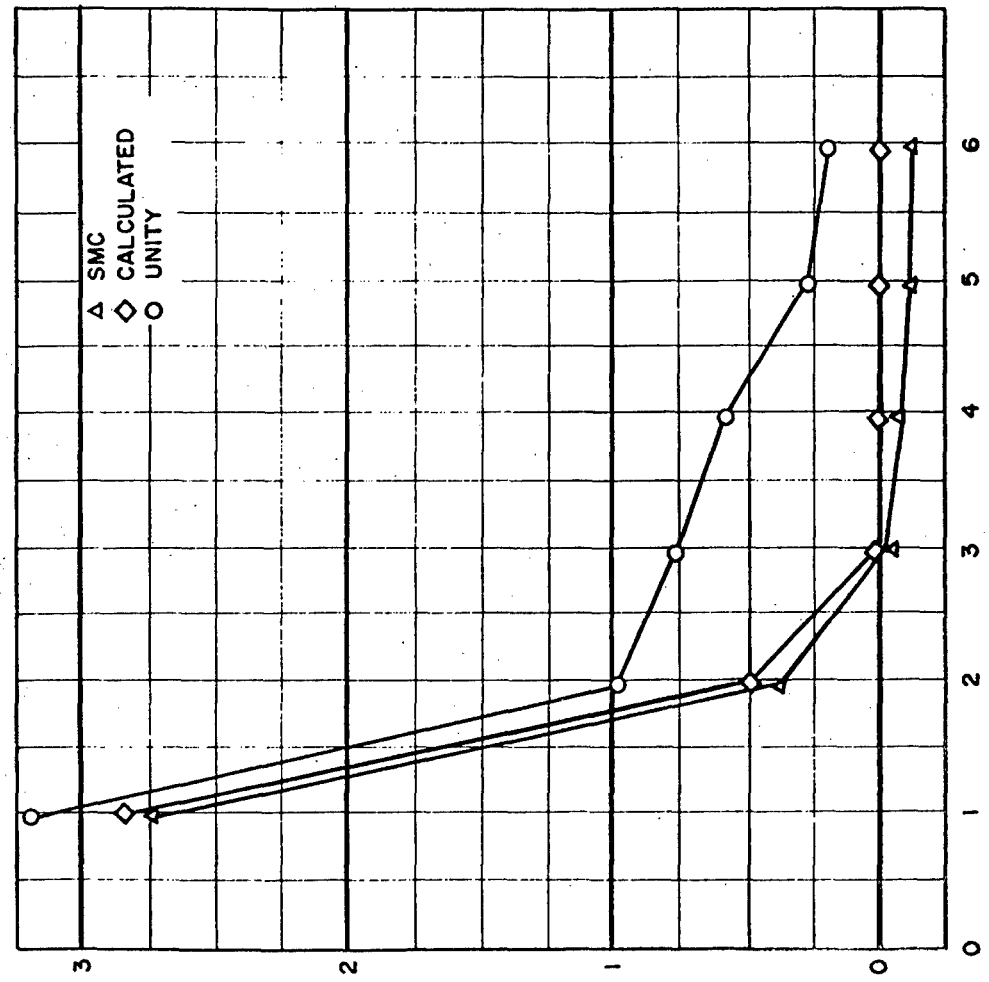


Figure 3. Eigenvalues of 6x6 Matrix

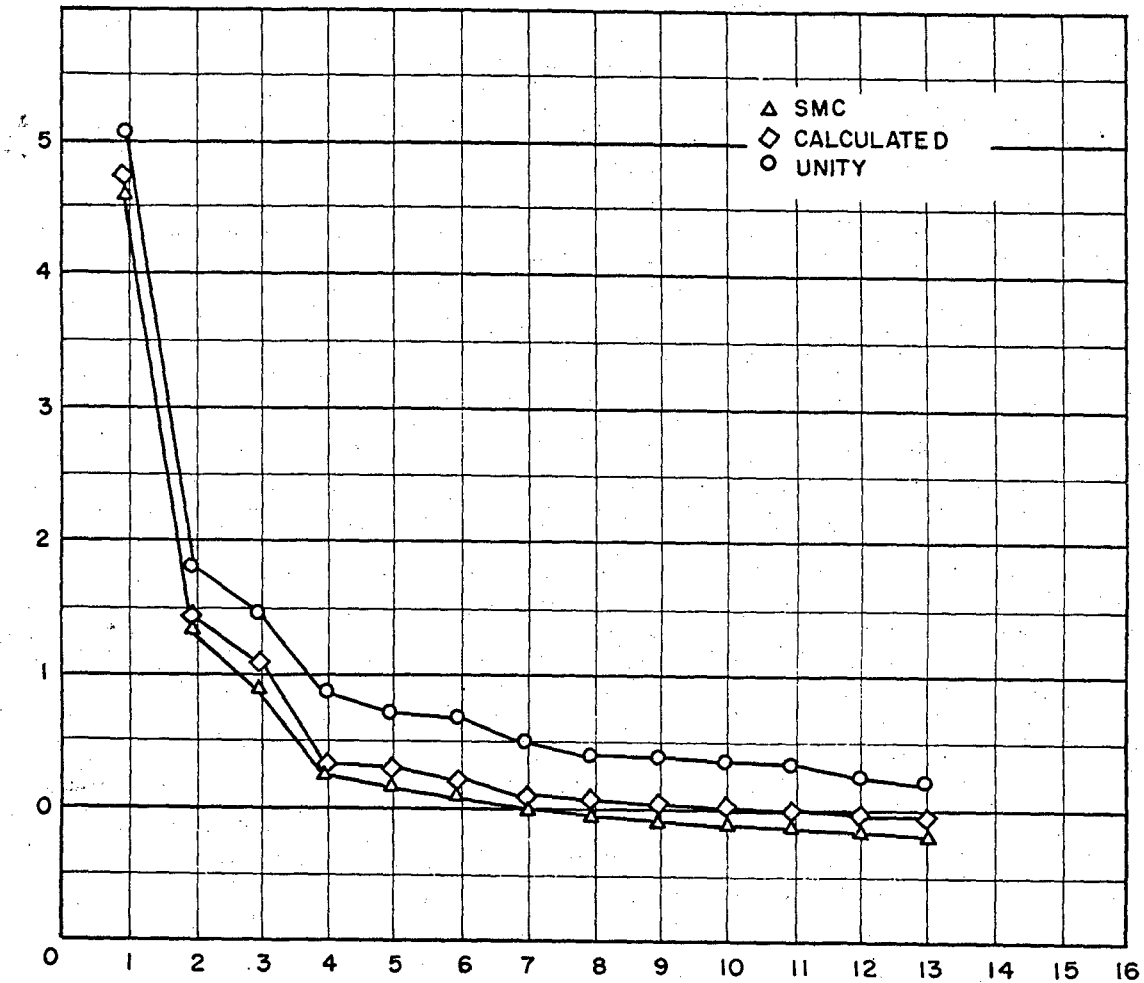


Figure 4. Eigenvalues of 13×13 Matrix

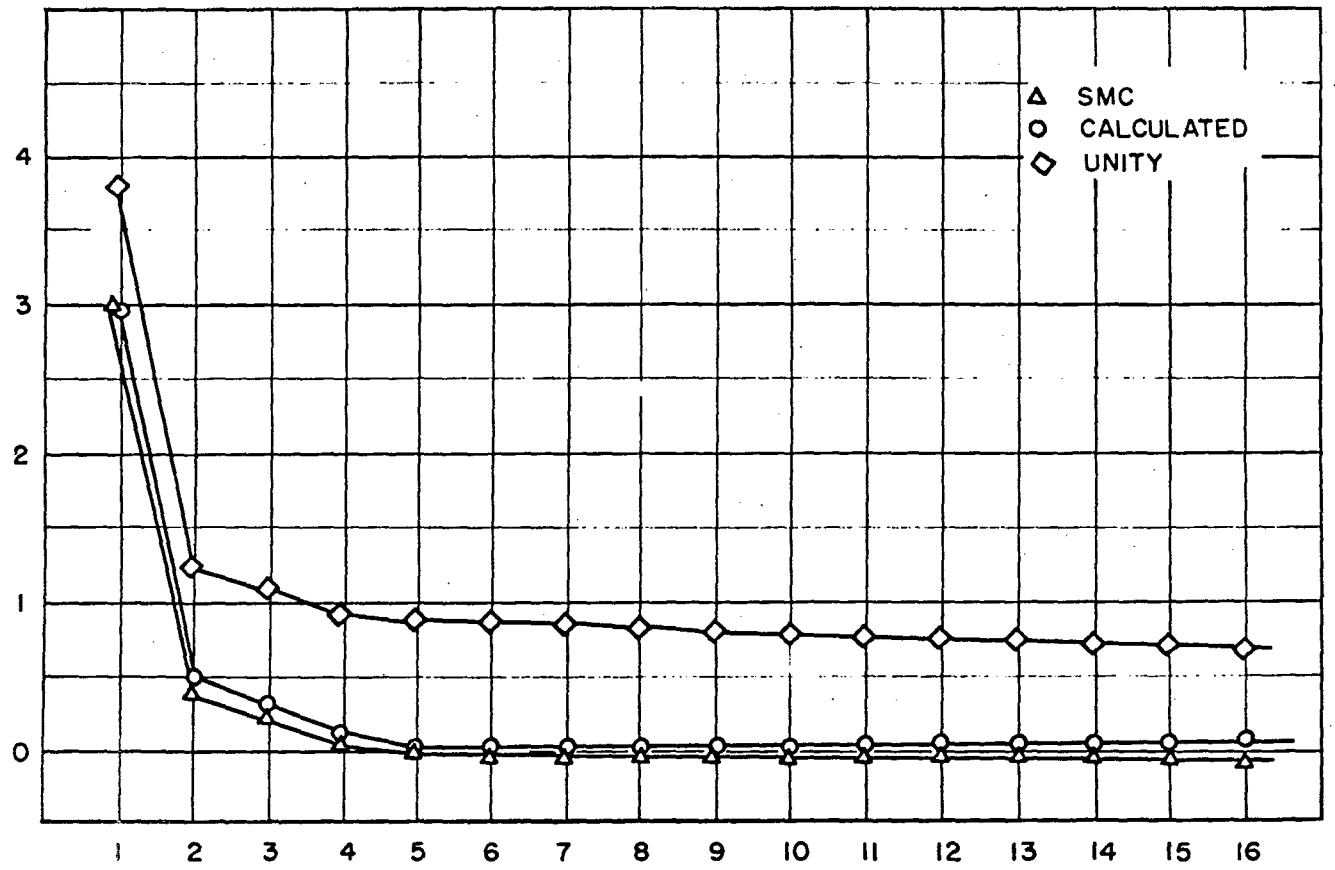


Figure 5. Eigenvalues of the 16×16 Matrix

C. Conclusions

The proposed method found diagonal elements which satisfied the refined definition of communality in all cases where an exact reduced rank was known and possible. In the other cases, the method found diagonal elements which satisfied the definition better (from the eigenvalue point of view) than did either squared multiple correlations or unities.

The method converged to "true communality" when either squared multiple correlations or unities were placed on the diagonal initially. However, the process converged faster with squared multiple correlations as initial values.

From a study of eigenvalue plots in several cases, it would appear that squared multiple correlation is a very good estimate of communality when there are only a few well-defined major factors. That is, either estimates of communalities are calculated by the method presented here using unities or squared multiple correlations as initial values for the method, or squared multiple correlations are themselves used as estimates of communalities. When factor analysis is used for the purpose of interpretation, the factor loadings are used to indicate which variables to associate with which factors. And the sets of associations are the same whether the factor loadings are obtained from a final reduced correlation matrix with communalities on the diagonal or squared multiple correlations on the diagonal. Thus squared multiple correlations are sufficiently close to true communalities to distinguish major factors when they exist.

4.5 COMPLETENESS OF FACTORIZATION

In factoring a correlation matrix no unique test as an answer to the question "when to stop factoring?" has yet been developed. At present there exist several methods which are applied with more or less success.

A few comparative or survey studies of some of these methods are available: Mosier (Reference 33) studies six different tests for completeness of factorization, applying them to one correlation matrix. Cattell (Reference 24) lists and evaluates eleven tests. Burt (Reference 34) summarizes, under the topic of "tests of significance in factor analysis" many of the existing methods. Fruchter (Reference 35) comparatively evaluates various tests, applying them to one or more concrete cases. In the most recent survey paper Sokal (Reference 36) evaluates comparatively five tests applying them to

Thurstone's box measurements, an artificial correlation matrix, a psychological matrix and a biological matrix, using centroid factor extraction. The present investigation compiles methods, which are used by factor analysts, in the form of a quick reference, listed in a systematic way.

Before leaving this introductory part let us make two remarks:

(a) Cattell (Reference 24) and also Burt (Reference 34) and Fruchter (Reference 35) suggest, that if one wants to rotate, it pays off to extract one or two more factors than necessary after application of any of the completeness tests, since one obtains more accurate results. Several workers also suggest applying more than one criterion and deciding upon the number of factors on the basis of the results of all the criteria.

(b) Obviously a solution to the communality problem together with the simultaneous knowledge of the rank will also resolve the completeness problem. The technique described in 4.4 presents such a solution. Since it is a converging process the adequacy of the factor solution of the original correlation matrix may then be shown by any of the following tests. It is suggested to then use one of the statistical tests, in order not to bring an empirically approximate view into the mathematically sound picture of the applied method.

For reference let us set up the following list of methods to test completeness of factorization:

A. Empirical completeness tests
   1. Percentage tests
   2. Tucker's test
   3. Cattell's scree-test
   4. Kaiser's test

B. Significance tests for completeness
   1. Tests for joint significance of residuals
      a. McNemar's test
      b. Saunder's test
   2. Tests for individual significance of residuals
      a. Test by means of standard error formula for the final residuals
      b. Sokal's test

3. Burt's chi-squared test

4. Lawley's chi-squared test

C. Miscellaneous tests for completeness

1. Index of completeness of factorization

2. Listing of other completeness tests

A. Empirical Completeness Tests

1. Percentage Tests

A practical and commonly used test for completeness of factorization considers percentages of total communality, accounted for by the factors. The tests can be conducted under different aspects:

(1) Determine in advance to analyze up to, say, 50% of the total variance, or a suitable proportion of the total reliability (leaving a proportion for the specificity).

(2) Determine in advance that a factor which accounts for less than, say, 5% of the total variance will not have any practical significance in the sense of being identifiable.

(3) Extract factors and if, after, say, 90% of the total communality or total variance have been accounted for, a factor accounts for only 2% of these totals, do not retain it in the set of factors.

The percentage tests are especially handy for the principal factor solution since the contribution of the factors to the total variance or total communality decreases with each succeedingly extracted factor. One could then stop factoring after one reaches a factor which accounts for, say, 5% of these totals. One knows that the next factor which could be extracted, would contribute less than 5% to the totals.

There is one more simplifying aspect of the principal factor solution. The total contribution

$$\sum_{j=1}^{n} a_{jp}^2$$

of factor F_p to the total variance or total communality, which is equal to the trace of the determinant of the correlation matrix, is equal to the λ_p-eigenvalue. The effect of each factor contribution to these totals

can therefore be computed easily as eigenvalue-percentage of the trace.

2. Tucker's Test

Denote by $\sum |\rho_k|$ the sum of absolute residuals of the $n \times n$ correlation matrix after k factors have been extracted.

Tucker's test (Reference 37):

If

$$\frac{\sqrt{\sum |\rho_{k+1}}}{\sqrt{\sum |\rho_k|}} \geq \frac{n-1}{n},$$

then the (k+1)-factor is considered to be insignificant.

Tucker's criterion after a modification by Blakey (Reference 38):

If

$$\frac{\sum |\rho_{k+1}|}{\sum |\rho_k|} \geq \frac{n-1}{n+1},$$

then the (k+1)-factor is considered to be insignificant.

Remarks:

(a) $\sum |\rho_k|$ and $\sum |\rho_{k+1}|$ include the communality residuals. Sokal (Reference 36) states, that it is desirable to use re-estimated communalities in place of residual ones in the denominator; but since the difference between residual and re-estimated diagonal values is usually slight, it is not of great importance what values are used in the main diagonal.

(b) Cattell (Reference 24) considers Tucker's test as one of the most reliable and practical ones of the really quick tests of completeness, though it sometimes can give strange results, since

the value of the ratio can decrease or increase after extraction of factors instead of increasing steadily.

Sokal, using the second form of the criterion in a comparative study of five tests for completeness of factorization, considers the test as not very suitable as a strong and fast criterion. Empirical investigations by McNemar (Reference 39) and theoretical investigations by Burt (Reference 34) support his standpoint. Burt criticized Tucker's test as making no allowance for the number of variables and the number of factors extracted and as making no explicit reference to the size of the sample. He considers the test as marking too many factors as insignificant.

Tucker's criterion has actually been employed by more factorists than any other criterion.

3. Cattell's Scree-test (for a principal factor solution)

Starting with the largest, each eigenvalue is plotted in an x-y-coordinate system, its number versus its length. Then the curve through these points is examined. If the number of factors, m, is less than the number of variables, n, n-m eigenvalues of the correlation matrix will be zero or at least close to zero, lying on a straight line almost parallel to the x-axis. The test consists in determining that point, where the curve breaks off the straight line. The number of eigenvalues determining the left part of the curve yields the number m of factors.

4. Kaiser's Test (Reference 40)

Upon extensive studies of correlation matrices with unities in the main diagonal Kaiser suggests as a practical basis for determining the number of common factors the number equal to the number of eigenvalues greater than one. Kaiser found that this number amounts to about a sixth or a third of the total number of variables.

B. Significance Tests for Completeness

1. Tests for Joint Significance of Residuals

a. McNemar's test

Let σ_k denote the observed standard deviation of the residuals (disregarding diagonal values) after extraction of k

factors. Let M_{h2} denote the mean of the communalities (computed from k factors). Then $1 - M_{h2}$ is the average uniqueness. If

$$\sigma_{res} = \frac{\sigma_k}{1 - M_{h2}} \leq \frac{1}{\sqrt{N}} ,$$

N being the number of observations, all significant factors have been extracted.

Remarks:

(a) McNemar's criterion is an attempt to test the significance of the residuals after k factors have been extracted from the correlation matrix. He used the centroid solution for his derivations. In the beginning years of factor analysis an attempt to do so was made by comparing the standard deviation of the residual correlations with the standard error of the original correlations. This device, though, is not adequate since residual correlations are analogous to partial correlations (the factors being held constant) and should for this purpose be divided by the geometric mean of the uniquenesses of the two variables under consideration. To reach his goal to test the significance of the residuals after k factors have been removed from the correlation matrix, McNemar approximates the standard deviation of the residuals or partial correlations by

$$\frac{\sigma_k}{1 - M_{h2}}$$

Cattell (Reference 24) reasons on the basis of experience that McNemar's test tends to stop factorization too early. Sokal (Reference 36) concludes from his studies that McNemar's test yields interpretable results except for problems with very large sample size N and low uniquenesses (that is, high communalities), in these cases indicating more than the true number of factors. In this respect it is worth noting that the test mainly takes into account the sample size N.

(b) Burt (Reference 34) suggests along the same line a procedure, which, as he says, is more satisfactory by not using residuals but

converted residuals (see Sokal's test) squaring them and summing them up. Then, if N is large, this sum will be approximately distributed as chi-squared. So, he suggests to determine significance of any particular set of residuals by referring to the χ^2-table with $\frac{1}{2}n(n-1) - kn + \frac{1}{2}k(k-1)$ degrees of freedom.

b. Saunder's test

Let $\sum\rho_k$ denote the sum of the residuals of the $n \times n$ correlation matrix after k factors have been extracted. Let N be the sample size and denote by

$$\sum_{j=1}^{n}\sum_{i=1}^{k} a_{ji}^2$$

the sum of all $n \times k$ factor loadings, the loadings taken from the unrotated matrix. The test can take on two forms: If

$$\sum\rho_k^2 \leq \left(\frac{n-k}{n}\right)^2 \frac{\left(n - \sum_{j=1}^{n}\sum_{i=1}^{k} a_{ji}^2\right)^2}{N}$$

after the k^{th} factor has been computed, then the factor extraction is complete. If the reliability coefficients of the variables are denoted by r_{jJ} then the test can be stated as: If

$$\sum \rho_k^2 < \left(\frac{n-k}{n}\right)^2 \frac{\left(n - \sum_{j=1}^{n} r_{jJ}\right)^2}{N}$$

after the k^{th} factor has been computed, then the factor extraction is complete.

Remarks:

(a) It is advisable not to include the diagonal residuals in $\sum \rho_k^2$ unless one is sure of exceptionally good communality estimates. If the communality residuals are excluded from the summation one has to multiply $\sum \rho_k^2$ by $\frac{2n}{n-1}$ to bring it to that equivalent with a whole matrix.

(b) Saunder's (Reference 41) claims his formula is an improvement over McNemar's test since it takes into account the sample size, the number of variables, the reliabilities and especially the number of factors.

(c) Sokal (Reference 36) applying Saunder's test to his four matrices obtains results similar to those obtained by McNemar's criterion. He again finds the apparent influences of large sample sizes or high communalities on the results.

2. Tests for Individual Significance of Residuals.

a. Test by means of standard error formula for the final residuals

Two approximate standard error formulas can be employed to decide upon the significance or insignificance of any residual after any number of factors has been extracted from the original correlation matrix.

(1) Theoretically it should be: $R = AA'$. Extracting common factors, R will only be reproduced by AA' approximately. How good this approximation is, or in other words, how complete factorization is, can be judged on the basis of the residual matrix $\bar{R}$, $R = \bar{R} + AA'$. Each element of $\bar{R}$, the final residual correlations, should be approximately zero in size; since, when all common factors have been extracted, no further correlation should exist between the variables. Let us assume, therefore, that the distribution of the residuals is similar to that of a zero-correlation in a sample of equal size. Then denote by $\sigma_{\bar{r}}$ the standard deviation of the series of residuals and by $\sigma_{r=0}$ the standard error of a zero-correlation. Under the above assumptions it would then be necessary as a test for completeness to determine if

$$\sigma_{\bar{r}} \leq \sigma_{r=0} = \frac{1}{\sqrt{N-1}}$$

or, since N is usually large, if

$$\sigma_{\bar{r}} \leq \frac{1}{\sqrt{N}}$$

From the application of this test-depending on the sample size alone, which is rather crude-one may conclude: If

$$\sigma_{\bar{r}} > \frac{1}{\sqrt{N}}$$

to an appreciable extent: further linkages between variables may exist; further factorization may be necessary. If

$$\sigma_{\bar{r}} < \frac{1}{\sqrt{N}}$$

to an appreciable extent: unjustified linkages between variables were factorized.

The above test can be found in Holzinger and Harman (Reference 42) and in Harman (Reference 2). Applications can also be found in these texts. Similar formulae have been proposed by Kelley (Reference 43) and Thurstone (Reference 44).

(2) Holzinger and Harman (Reference 42) have derived a standard error formula for a residual after any number of factors has been extracted from the correlation matrix.

Denote by $\bar{r}_{ij}$ the observed correlation between variable i and j, by $\bar{r}_{ij}$ the residual after extraction of $m+1$ factors, by $\sigma^2_{a_{is}}$ and $\sigma^2_{a_{js}}$ the standard errors of factor loadings, then

$$\sigma^2_{r_{ij}} = \sigma^2_{r_{ij}} + \sum_{s=0}^{m} \left(a^2_{is} \sigma^2_{a_{js}} + a^2_{js} \sigma a^2_{is} \right)$$

This formula, however, cannot be applied to a residual obtained from any solution since the standard errors $\sigma^2_{a_{js}}$ and $\sigma^2_{a_{is}}$ are only known for the two-factor and bi-factor solutions.

In approximating the above formula, so that it does not explicitly contain the standard errors of the loadings, the assumption is made that all observed correlations can be well enough described by their average, computed by

$$\rho = \frac{1}{2\binom{n}{2}} \sum_{i,j} (r_{ij} \; ; \; i,j = 1,\ldots, n, \; i \neq j) \; .$$

And if ρ_s denotes the average residual correlation used for computing loadings of the factor F_s, the approximate formula after extraction of $m+1$ factors is of the form:

$$\sigma_r^2 = \frac{(1-\rho)^2 (5 + 8\rho + 2\rho^2)}{2N} + \frac{1}{N}\sum_{s=1}^{m}\left(\frac{3}{2} - \rho_s - \frac{5}{2}\rho_s^2 + 2\rho_s^3\right) .$$

These standard errors are tabulated in References 42 and 2. Applications can also be found there.

It should be noted, that the necessary approximations to arrive at above formula, make the σ_r-value usually smaller, so, in order to take this fact into account, a residual which is twice its standard error can still be considered insignificantly different from zero.

b. Sokal's Test

In the following test each single residual is tested for insignificance. Denote by $\rho_{ij.k}$ the residual correlation between variables i and j after k factors have been extracted. Let u_{ik}^2 denote the uniqueness of variable i after extraction of k factors:

$$u_{ik}^2 = 1 - a_{i1}^2 - a_{i2}^2 - \cdots - a_{ik}^2 .$$

Convert the residuals to quantities analogous to partial correlations (factors through k kept constant) by dividing them through the geometric mean of the uniquenesses of the variables under consideration. Name the converted residuals $r_{ij.12\ldots k}$. So

$$r_{ij.12\ldots k} = \frac{\rho_{ij.k}}{\sqrt{u_{ik}^2\, u_{jk}^2}} = \frac{\rho_{ij.k}}{u_{ik}\, u_{jk}}$$

Assume that the $r_{ij.12\ldots k}$ have the same sampling distribution as ordinary partial correlation coefficients. Under this assumption test each converted residual against the minimum significant partial correlation, denoted by $r_{m.s}$, obtained from table IV, Fisher and Yates (Reference 45):

If

$$r_{ij.12...k} < r_{m.s}$$

at a presumed level of significance with N - (k+1) degrees of freedom, then $r_{ij.12...k}$ is insignificant.

Remarks

(a) The rather laborious work to conduct the test on each residual can be simplified by excluding certain residuals from the test. This is done by the following procedure: for a presumed significance level, $r^2_{m.s}$ can be determined as well as the lowest two uniquenesses, denoted by $_m u^2_{ik}$ and $_m u^2_{jk}$. Then from

$$r^2_{m.s} = \frac{\rho^2_{ij.k}}{{}_m u^2_{ik}\ {}_m u^2_{jk}}$$

$\rho_{ij.k}$ can be determined:

$$\rho^2_{ij.k} = r^2_{m.s}\ {}_m u^2_{ik}\ {}_m u^2_{jk} = \rho^2_{m.s}$$

and all values

$$\rho_{ij.k} < \rho_{m.s}$$

are certainly unsignificant. That means, for the test only values

$$\rho_{ij.k} > \rho_{m.s}$$

have to be considered.

(b) Sokal (Reference 36) discusses this completeness test in his comparative study, mentioning also some computing simplifications. He obtains his results by judging the elements of the residual matrix by the

above described significance test and by the "importance" test, that is by counting the number of partial correlations larger than an arbitrary 0.05 (disregarding the sign), that remain in the matrix after k factors have been extracted, naming those correlations important. On the basis of his study he recommends these procedures to test for completeness because of the statistical basis of the significance test and the apparent consistent results.

(c) In one of his early tests Burt (Reference 46) started from the same considerations as Sokal, defining $r_{ij.12...k}$ and testing it against the standard error of a zero partial correlation, $\frac{1}{\sqrt{N}}$, then the test is given by

$$r_{ij.12...k} = \frac{\rho_{ij.k}}{u_{ik}\, u_{jk}} \leq \frac{1}{\sqrt{N}} .$$

3. Burt's Chi-squared Test with Z-transformation

Theoretically, it is $R = AA'$. Test the significance of the differences between the elements of R and of AA' after k factors have been extracted. Let $\overline{Z}$ denote the elements of R transformed by Fisher's Z-transformation and let Z denote the elements of AA', also transformed by Fisher's Z. Sum $(Z-\overline{Z})^2$ over the upper or lower triangles (without diagonals) of the respective matrices. If N is the sample size, n the number of variables the test of significance is expressed by: If

$$\chi^2 = (N - 3) \sum (Z-\overline{Z})^2$$

with $\frac{1}{2}n\,(n - 1) - kn + \frac{1}{2}k\,(k - 1)$ degrees of freedom is insignificant at a presumed level of significance, the factor extraction is assumed to be completed.

Remarks:

(a) Fisher's Z-transformation, $Z=\tanh^{-1} r = \frac{1}{2}\log_e \frac{1+r}{1-r}$, is applied to the elements of R and AA' to obtain their normal distribution

(b) Burt recommends this test in his 1952 paper (Reference 39) as the "most useful available when current factorial procedures are employed".

Sokal (Reference 36) in his comparative study obtains some correct results and points out the fact that small correlation matrices may not provide enough degrees of freedom.

4. Lawley's Chi-squared Test

In the following we will consider a statistical test for the number of common factors. This test should be used, though, for large samples only and with ones in the main diagonal of the correlation matrix.

Let N denote the sample size, $|R|$ the determinant of the matrix of observed correlations and $|\hat{P}|$ the determinant of the maximum likelihood estimator ($\hat{P} = \hat{A}\hat{A}' + \hat{a}^2$ where factor loadings are determined by the maximum likelihood method) of the population correlation matrix. Let the variables have a multivariate normal distribution. Then

$$\text{chi-square} = \chi^2 = N \log \frac{|\hat{P}|}{|R|} \qquad (4)$$

with

$$\nu = \frac{1}{2}[(n-k)^2 - n - k]$$

degrees of freedom is used to test the hypothesis that k common factors adequately explain the correlations at an assumed level of significance.

Lawley (Reference 42), who derived the above formula, simplified it, by approximation to the following χ^2 formula to be examined:

$$\chi^2 = N \sum_{i<j=1}^{n} \frac{\bar{r}_{ij}^2}{a_i^2 a_j^2} \qquad (5) \text{ (corrected residuals)}$$

where $\bar{r}_{ij}$ denote the residuals obtained by

$$\bar{r}_{ij} = r_{ij} - r'_{ij}$$

with r'_{ij} being the elements of $\hat{P}$, that is the (maximum likelihood estimated) reproduced correlations.

Remarks

(a) Harman (Reference 2) states that, usually, that is by other than statistical means, one underestimates the number of statistically significant factors, compared with the number of factors one obtains by application of the χ^2test.

(b) As Harman points out, it is reasonable to apply the test also to problems where the maximum likelihood method is not employed to estimate $\hat{P}$ if one draws only a conclusion in the case where the χ^2value is found to be insignificant. In case the χ^2value is significant, though, no conclusion can be made since it is possible that a maximum likelihood factorization gives better results.

(c) Rippe (Reference 48) arrives at a formula identical with the likelihood ratio (equation 4), his development not being specifically dependent on maximum likelihood estimates of factor loadings.

(d) An experimental study of the test was furnished by Henrysson (Reference 49).

C. Miscellaneous Tests for Completeness.

1. Index of Completeness of Factorization

If the uniqueness a_j^2 of a variable X_j is broken down in unreliability c_j^2 and specificity b_j^2, that is

$$a_j^2 = b_j^2 + c_j^2 ,$$

then the index of completeness of factorization is defined by

$$H_j = \frac{100\ h_j^2}{h_j^2 - b_j^2} \qquad (h_j^2 \text{ the communality}).$$

This index can well be used to decide whether factorization was carried too far or not; for almost no variable X_j should H_j be in excess of 100. Especially in the analysis of psychological tests into common factors, this analysis should not be carried to the point where real specific factors disappear.

2. Listing of Other Completeness Tests

There does exist a variety of other methods for checking completeness

of factorization. Since they are partly simple ones from the early days of factor analysis or do not have important effects on factor analysis, we will indicate them here only, referring to the papers where they can be found.

(1) Plotting the distribution of the residuals after extraction of k factors and comparing this distribution with the normal curve is described in Cattell (Reference 24, pp. 297-298) as completeness check.

(2) See Mosier (Reference 33) for a comparison of six simple methods. A short description of three of these methods, which were found to be rather effective, is given in Cattell (Reference 24).

(3) See Reyburn and Taylor (Reference 50) for a method which compares the frequency distribution of the quotients of a residual over the standard error of its corresponding original correlation with the normal distribution.

(4) Coombs (Reference 51) suggests a test for the centroid solution by counting the number of negative signs left in the residual matrix after every possible variable reflection has been carried out and compares them with the number C of a table set up by Coombs, which depends on the number of variables.

(5) Swineford (Reference 52) correlates the original correlations with the series of corresponding residuals and continues factorization until this correlation becomes insignificant.

(6) Hoel (Reference 53) attempts in his paper, less fruitfully though, the development of a significance test for the number of common factors. See also Burt (Reference 34) for a short outline of the method.

(7) Wilson and Worcester (Reference 54) describe a chi-squared test.

(8) Young (Reference 55) derives an index of clustering.

(9) In the situation where we are dealing with component analysis (unities are employed in the main diagonal of the correlation matrix) Hotelling (Reference 56) and Bartlett (Reference 57) have provided statistical tests for the number of significant factors.

(10) Humphrey (see Fruchter, Reference 35) defined a completeness

criterion which takes into account the sample size and depends on the loadings of only two variables. He multiplies the two highest loadings in a column of the centroid factor matrix and compares the product with the standard error of the zero correlation coefficient to establish the significance or insignificance of the factor under consideration.

(11) It is noted for information that there exists a listing of twenty-five completeness criteria by Vernon, et al. (Reference 58).

4.6 EIGENVALUES AND THEIR BOUNDS.

A critical problem in factor analysis is the determination of the sample size, denoted by the number of observations N. This problem can be seen with respect to direct dependence of N on the number of variables or with respect to the factor analysis one wants to conduct. The question for the dependence of the number of observations on the number of variables is answered by factor analysts by such rules of thumb as: the ratio of the number of observations to the number of variables shall exceed 3 (or shall exceed 5); the number of observations minus the number of variables shall exceed 80. No good mathematical means has as yet been obtained for a better determination of this relationship. One indication of this relationship can be exhibited, however. On a geometrical basis (see 2.4) one finds that, if n = number of variables, m = supposed number of common factors and the factors are considered to be uncorrelated, then the m common factors and n unique factors are represented in N-space such that $m + n \leq N$, which determines: $N > m + n$.

The investigation reported in this subsection takes the second way of approach to the problem, namely to consider the sample size N in the light of the factor analysis to be conducted. In considering at all the problem of how large the sample should be, we are assuming, that if we would arbitrarily choose an N without reflecting upon anything, we might obtain less "reliable" factors. Here we want to understand by a reliable factor a factor whose loadings would change only little if the factor analysis would be conducted on a correlation matrix of the same variables but with a larger number of observations.

The solution to the problem was attempted to be found in statistical properties. Two assumptions had to be made: firstly, the assumption that all elements r_{jk} of the correlation matrix R be greater than 0, denoted by $R > 0$, and secondly, the assumption that the population of pairs (X_{ji}, X_{ki}), $i = 1,\ldots, N$, from a sample of which each element of R is computed, satisfies the bivariate normal distribution model. The first assumption is not so stringent, since many correlation matrices with small negative entries can be reduced to this form, the second assumption is one which is mostly made to guarantee statistical considerations on r_{jk}. The case, where some elements of R are equal to zero, can be considered also, if only R satisfies the "irreducibility" properties, which will be introduced a little later.

The statistical means to associate sample size N with the loadings of the factors, obtained by factor analyzing the correlation matrix R, is found in the confidence intervals, which one can compute for each element r_{jk} of R. By forming confidence intervals we assume that the observed correlation coefficients are only estimates of the true population correlation coefficients. The larger N is, the more does the observed coefficient approach the population coefficient, so that the difference between the observed r_{jk} and the confidence limits can be called the error due to N. Now we are interested in how these errors propagate through the factor analysis. Since the most popular method for obtaining a factor analysis of R is the principal-factor method, where the factor loadings are directly computed from eigenvalues of the correlation matrix, the question we ask is the following: How much does the error, introduced into the correlation matrix R by way of the fact that the elements of R are only N-dependent estimates of the true correlation coefficient, influence the eigenvalues of R?
To obtain information about this, the following procedure is suggested. For each r_{jk} confidence limits are computed according to the technique outlined in Section 2.5. For each r_{jk} we obtain two confidence limit-values, which we denote by $r_{jk}^{(1)}$ and $r_{jk}^{(2)}$ with $r_{jk}^{(1)} < r_{jk} < r_{jk}^{(2)}$
If an r_{jk} is computed to be insignificantly different from zero, we insert the value 0.001 (or if there is an $r_{jk} < 0.001$, an even smaller

value than 0.001) for it into R_1 (since we do not want any actual zero values in R_1); the value r_{jk} itself, however, is inserted in R_2. Expressing the above in matrix notation we obtain if $r_{jk} \in R$, $r_{jk}^{(1)} \in R$, and $r_{jk}^{(2)} \in R_2$: $R_1 < R \leq R_2$. Then we conduct principal-factor-analysis on the three matrices R_1, R, and R_2. We encounter some difficulties here. If we assume ones in the main diagonals of R, R_1, and R_2, then R is Gramian, while R_1 and R_2 are symmetrical but not necessarily positive semidefinite. On the other hand, seldom is a factor analysis done on R with ones in the main diagonal; rather squared multiple correlations or other communality estimates are inserted in the diagonal. So, also R differs slightly from being Gramian. How bad it is non-Gramian is determined by the number and size of negative eigenvalues. If they are small and few in number they can be neglected. We make use of this fact for the eigenvalues of R_1 and R_2. If N is large, R_1 and R_2 approximate R closely, so that they will not be too non-Gramian.

Under the assumption that $R > 0$ also $R_1 > 0$ and $R_2 > 0$. This is based on the fact that the confidence intervals for each element of R do not exceed over the zero point. If they would exceed over the zero point, the population correlation coefficient could be zero. But this is excluded from consideration since each correlation coefficient is first tested for this hypothesis and the confidence limits are only computed if the population coefficient is not equal to zero.

Thus, since $R_1 < R \leq R_2$ and $R_1 > 0$, $R_2 > 0$ we can express R_1 and R_2 as: $R_1 = R - E_1$, and $R_2 = R + E_2$, respectively, where E_1 has only positive entries and E_2 has positive and (or only) zero entries.

Our objective will now be to show the following: If $r_{jk}^{(1)}$ and $r_{jk}^{(2)}$ represent the lower and upper 95%-confidence limits on the correlation coefficient r_{jk}, by having defined the $r_{jk}^{(1)}$ and $r_{jk}^{(2)}$ values if r_{jk} is insignificant as above, such that $r_{jk}^{(1)} < r_{jk} \leq r_{jk}^{(2)}$, and if $r_{jk}^{(1)} \in R_1 > 0$, $r_{jk} \in R > 0$, and $r_{jk}^{(2)} \in R_2 > 0$, then $\lambda_1 < \lambda \leq \lambda_2$, where λ_1, λ, and λ_2 are the largest eigenvalues obtained

for the matrices R_1, R, and R_2.

Since the loadings of the first factor are directly computed from the largest eigenvalue, the result, which we will prove below, clearly links the loadings (by which we judge a factor analysis to be reliable) to the sample size N: the larger N is, the smaller will the interval $(r_{jk}^{(1)}, r_{jk}^{(2)})$ be, and correspondingly the interval (λ_1, λ_2).

Now let us prove the statement $\lambda_1 < \lambda < \lambda_2$ (under the above made assumptions). As we pointed out earlier in this subsection, we can make the assumption $R \geq 0$, but then R has to satisfy the irreducibility condition introduced by the following

Definition 4.1: For $n \geq 2$ an $n \times n$ matrix R with real elements is called reducible if there exists an $n \times n$ permutation matrix P (defined as a square matrix which in each row and in each column has some one entry unity, all others zero), such that

$$PRP^T = \begin{pmatrix} R_{1,1} & R_{1,2} \\ 0 & R_{2,2} \end{pmatrix}$$

where $R_{1,1}$ is an $r \times r$ submatrix and $R_{2,2}$ is an $(n - r) \times (n - r)$ submatrix with $1 \leq r < n$. If no such permutation matrix exists, then R is called irreducible. If R is a 1×1 matrix, then R is irreducible if its single entry is nonzero and reducible otherwise.

In the proof of our statement we will have to use either one of two theorems, according to the assumptions made on R. If $R > 0$, we shall use Perron's Theorem (Theorem 3.8), if $R \geq 0$ and R is irreducible we shall use the following Theorem 4.4, due to Froebenius, an extension of Perron's Theorem to irreducible matrices.

Theorem 4.4: An irreducible matrix $R \geq 0$ always has a positive eigenvalue λ which is a simple root of the characteristic equation. The moduli of all other characteristic numbers are at most λ. The eigenvector corresponding to λ has positive components and is essentially unique (up to scale factors).

The proof of Froebenius' Theorem can also be found in Gantmacher (Reference 22).

Proof of the statement $\lambda < \lambda \leq \lambda_2$, where λ_1 is the largest eigenvalue of R_1, λ the largest of R, and λ_2 is the largest of R_2. If x is an eigenvector belonging to λ, and x_1 is an eigenvector belonging to λ_1 we have

$$Rx = \lambda x \tag{6}$$

$$R_1 x_1 = \lambda_1 x_1 \tag{7}$$

We have

$$R_1 = R - E_1 , \quad E_1 > 0,$$

and taking the inner product of Equation 7 with x we obtain:

$$(x|R_1 x_1) = \lambda_1 (x|x_1)$$

$$(R_1 x|x_1) = \lambda_1 (x|x_1) \quad \text{since } R_1 \text{ is real and symmetric}$$

$$\left|(R - E_1)\, x|x_1\right| = \lambda_1 (x|x_1)$$

$$(Rx|x_1) - (E_1 x|x_1) = \lambda_1 (x|x_1)$$

$$\lambda (x|x_1) - (E_1 x|x_1) = \lambda_1 (x|x_1)$$

$$\lambda - \frac{(E_1 x|x_1)}{(x|x_1)} = \lambda_1$$

$(x|x_1)$ is not equal to zero, since x and x_1 have only positive

components, according to Perron's (or in case it is applicable, Froebenius') Theorem, applied to R and R_1. So the term $(E_1x|x_1)/(x|x_1)$ is positive and thus $\lambda_1 < \lambda$. In the same manner it is proved that $\lambda \leq \lambda_2$, with the equality holding if E_2 is the zero matrix, which makes $(E_2x|x_2)/(x|x_2)$ equal to zero.

Some remarks about the result shall be made next. From the analysis of the principal-factor method it follows that:

$$\lambda_1 = \sum_{j=1}^{n} {}_{(1)}a_{j1}^2 ,$$

$$\lambda = \sum_{j=1}^{n} a_{j1}^2 ,$$

and

$$\lambda_2 = \sum_{j=1}^{n} {}_{(2)}a_{j1}^2$$

or, the largest eigenvalues of R_1, R, and R_2 are equal to the sum of the contributions of the first factor (in each respective factor analysis) to the total communality of each analysis. The length of the interval for λ, namely the difference

$$\lambda_2 - \lambda_1 = \sum_{j=1}^{n} {}_{(2)}a_{j1}^2 - \sum_{j=1}^{n} {}_{(1)}a_{j1}^2 = \sum_{j=1}^{n} \left({}_{(2)}a_{j1}^2 - {}_{(1)}a_{j1}^2 \right)$$

is the largest difference which we can get between the sum of the squared factor loadings of the two first factors, obtained by factor analyzing R_1 and R_2. The difference approaches zero when N increases, since the length of the interval for λ then becomes smaller.

It was wished to determine the sample size N. The difference

$$\sum_{j=1}^{n} {}_{(2)}a_{j1}^2 - {}_{(1)}a_{j1}^2$$

indicates how much the contribution of the first factor can vary in

dependence on N. In other words, it can be checked whether an assumed N is large enough so that the variation of the first factor contribution to the total communality of R does not exceed a given value (which could perhaps be computed as a percentage of the total communality).

Example: An example was computed to show the proposed method. The problem of 24 psychological variables, whose correlation matrix and analysis are reported in Harman (Reference 2, page 137 and page 185) was taken for this example. The one insignificant negative value, which Harman's matrix contains, was changed to an insignificant positive one in order to meet the requirements for application of Perron's Theorem. The sample size as given in Harman is N=145.

Let us briefly outline the kind of computations done for the example.

(1) The two matrixes R_1 and R_2 were computed according to the discussion in this subsection. The value 0.001 was inserted into R_1, while the values r_{jk} themselves were inserted into R_2 when r_{jk} was found to be insignificant.

(2) Squared multiple correlation coefficients were computed for the three matrices.

(3) Factor analyses were conducted on the matrices R_1, R, and R_2. The eigenvalues and factor loadings were obtained.

For a comparative study let us now consider the obtained values. We list the postive eigenvalues in Table 4 and then the first-factor loadings, computed from the 3 first (largest) eigenvalues in Table 5. It is also interesting to list the following data:

	Total Original Communality	Sum of Positive Eigenvalues	Sum of Negative Eigenvalues
R_1	7.9184	9.8388	1.9204
R	11.8761	13.4935	1.6174
R_2	21.6238	22.0108	0.3870

Table 4

The Positive Eigenvalues of R_1, R, and R_2

R_1-eigenvalues	R-eigenvalues	R_2-eigenvalues
λ_1 = 4.3884	λ = 7.6665	λ_2 = 10.8149
1.6844	1.6634	2.2210
1.1014	1.1785	1.6392
0.8292	0.9212	1.4155
0.4342	0.4319	0.9245
0.3608	0.4064	0.8684
0.2811	0.3199	0.6861
0.2512	0.3024	0.6596
0.2136	0.2513	0.5467
0.1482	0.1759	0.4625
0.1084	0.1082	0.3760
0.0379	0.0433	0.3235
	0.0246	0.2860
		0.2108
		0.1943
		0.1598
		0.1358
		0.0576
		0.0286

Interpreting the obtained results, the following can be said:

(1) The 3 matrices can be considered as not too non-Gramian, the size of the negative eigenvalues being small. Especially, the number and size of the negative eigenvalues of R_2 are small. Here, though, a difficulty arose when a squared multiple correlation coefficient, as estimate of communality, turned out to be larger than one (based on the fact that R_2 with ones in the main diagonal has not, as R does, the representation as $R = ZZ^T/N$).

(2) Table 5 shows the expected results that all factor loadings of the three first factors--as derived from positive eigenvalues and eigenvectors--are positive.

(3) It is interesting to note, that both R_1 and R show four distinctively large eigenvalues while there is a sharp drop in the size of the eigenvalues after the fourth ones. Each time the four eigenvalues account for more than 95% of the total original communality Harman suggests the interpretation of four factors, which is applicable to the results of R_1 . R_2 shows six distinctive eigenvalues with

Table 5

The First-Factor Loadings Computed from the Three First Eigenvalues

$\lambda_1 = 4.3884$	$\lambda = 7.6665$	$\lambda_2 = 10.8149$
factor loadings	factor loadings	factor loadings
0.4293	0.5952	0.7012
0.1594	0.3751	0.4784
0.2222	0.4297	0.5637
0.3130	0.4839	0.5906
0.6337	0.6901	0.7729
0.6336	0.6883	0.7620
0.6291	0.6728	0.7407
0.5930	0.6819	0.7595
0.6563	0.6898	0.7540
0.2744	0.4649	0.5586
0.3711	0.5588	0.6725
0.2643	0.4669	0.5873
0.4308	0.6038	0.7181
0.2016	0.4268	0.5544
0.1465	0.3896	0.5334
0.2897	0.5144	0.6395
0.2327	0.4631	0,6180
0.2877	0.5177	0.6614
0.2284	0.4511	0.5702
0.4831	0.6164	0.7288
0.4252	0.5969	0.7285
0.4746	0.6129	0.7229
0.5827	0.6895	0.7876
0.5174	0.6532	0.7621

four of them being over one. But one has to consider 11 eigenvalues to account for 95% of the original total communality, while 4 (6) eigenvalues account for slightly more than 70% (80%) of the original total communality.

(4) As for the main objective, the determination of N, the result shows that N = 145 is too small to furnish a reliable factor analysis. Already the confidence intervals are very large. For example:

$$0.013 < 0.176 < 0.330 \quad \text{for a small } r_{jk}$$

$$0.635 < 0.723 < 0.793 \quad \text{for a large } r_{jk}$$

The interval on the first eigenvalue is consequently also large: $4.3884 < 7.6665 < 10.8149$, so that the difference

$$\lambda_2 - \lambda_1 = 10.8149 - 4.3884 = 6.4265$$

is even not expressible as a reasonably small percentage of the original total communality of R.

Disregarding R_2 (for its difficulties of obtaining communalities larger than one) and considering only R and R_1 we compute $\lambda - \lambda_1 = 3.2781$ which is 27.6% of the original total communality 11.8761, still considerably high.

It must be concluded, that the sample size N = 145 is too small and it would be desirable to have more observations and to do the factor analysis over. On the other hand, both the R_I and R-analysis yield the same number of factors used for interpretation, which might suggest the contrary. This emphasizes the fact, which also Harman indicates, that proper statistical considerations are often lengthy but do not furnish better results.

4.7 FACTOR SCORES

The computational problem of representing observed variables in terms of hypothetical variables or factors F is only partly solved when the factor loadings A are computed. The factor loadings serve to describe the number of factors and the saturation of variables by a factor. And for some purposes, such as interpretation of factors, the loadings are sufficient. However, the complete representation is obtained only when the factors themselves are also computed.

In the case where the factor pattern takes the form

$$Z = AF$$

due to inserting unities on the diagonal of the correlation matrix, i.e., no unique factors are allowed or postulated, the common factors F may be solved for directly since the matrix A is a square $n \times n$ nonsingular matrix. Indeed

$$F = A^{-1}Z \quad .$$

When communalities are placed on the diagonal of the correlation matrix, the number of common and unique factors is greater than the number of variables, and therefore the factor loading matrix is singular with no inverse. In keeping with the original assumption of factor analysis that each variable is a linear function of the factors, it is now assumed that each factor is a linear function of the variables. However, since there are more factors than variables, the factors defined by the original linear form can only be estimated in a least squares sense by the linear form,

$$\overline{F}_p = \beta_{p1}Z_1 + \cdots + \beta_{pn}Z_n \quad (p = 1,2,\ldots,m) \quad .$$

It is shown (Reference 2, p. 340) that

$$\overline{F}_p = S_p^T R^{-1} Z \tag{8}$$

gives least squares estimates of the factors, where the subscripts denote columns. Factors estimated using Equation 8 have zero mean and a standard deviation close to one but varying from factor to factor.

Section V

THE ROTATION PROBLEM

5.1 INTRODUCTION

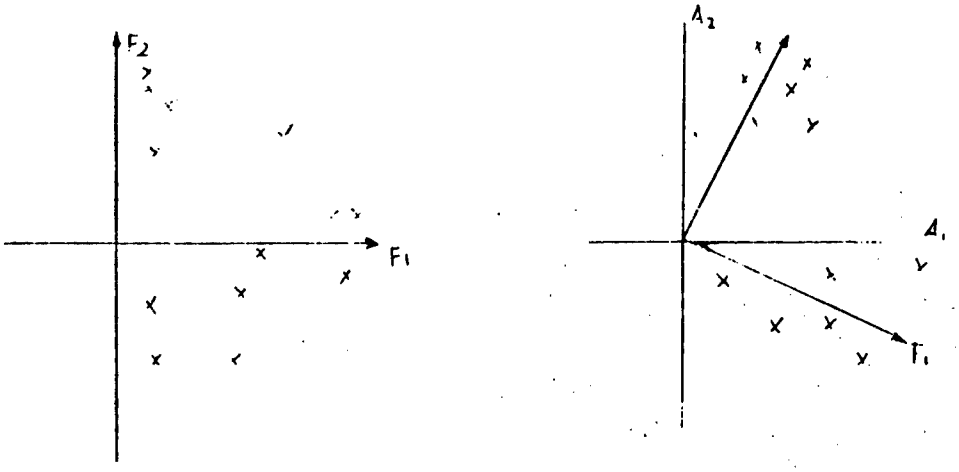
After we have discussed factor solutions and problems pertaining to them the next step in factor analysis is rotation. This problem is therefore considered in this section. In 5.2 the rotation problem will be stated. 5.3 gives a survey of existing rotation techniques. A specially interesting problem is the problem of interpreting oblique factors. Many factor analysts prefer to keep to orthogonality since the problems, raised by the fact that in the oblique rotation, factor pattern and factor structure are no longer equal, cannot satisfactorily be taken care of. On the other hand, an oblique solution might be the only adequate solution to a given problem. Therefore the important topic of interpretation of oblique factors is taken up in 5.4.

5.2 THE ROTATION PROBLEM

The second part of every worthwhile factor analysis is factor rotation. This procedure involves accepting a factor pattern (and other matrices in the oblique case) with an already determined number of factors and performing sequence of iterative matrix operations on it to re-orient the factor reference frame according to preset boundary conditions or constraints. The basic correlation matrix with communalities is preserved and must still be the result of FF^T (in model form).

There is an infinite number of ways to rotate the primary factor pattern which results from, say, a centroid or principal components analysis. Consider for a moment the analogy of defining the loci of points equidistant from the origin of a Cartesian coordinate system, each point simply representing another orientation of the end of line in 2-space. A graphical illustration of a typical two-factor rotation where the variables are represented by points in a plane is shown below.

The rotation of the reference frame to a "preferred" or "Simplified" position is both difficult and ambiguous. It is this process which is the cause of much controversy concerning the definition of a preferred, simplified or best solution. There have been and still are several schools of thought on this issue dating back to the origin of factor



analysis. The most popular definition is one developed by the psychologists and is used extensively today--simple structure. Other structures which are used from time to time include multiple-group, uni-factor, and bi-factor and are characterized by a preset factor pattern into which the loadings are to be fitted. Simple structure, on the other hand, represents a quasi-definite ordering of a desirable multiple-factor solution based on factor interaction experience of the behavorial scientists. The resultant pattern initially was one containing mostly very high and very low loadings distributed in such a way that the following three conditions were met:

1. each row should contain at least one zero
2. each column should contain at least as many zeros as there are common factors
3. for every pair of factors there should be at least m variables which do not load high on both factors (m being the number of common factors)

These conditions were first established by Thurstone (Reference 44) and later extended by him to provide "insurance"in his own studies as follows (Reference 25):

1. each row should contain at least one zero

2. each column should contain at least as many zeros as there are common factors

3. for every pair of factors there should be several variables which do not load high on both factors

4. for every pair of factors a large proportion of variables should have zero loadings on both factors when there are more than three factors

5. for every pair of factors there should be only a small number of variables with nonzero loadings on both factors.

If simple structure s decided to be the acceptable format for a factor pattern, one may choose from several factor rotation techniques, each of which provides a slightly different variation of the main theme. If another factor structure is desired, rotation may be exceedingly complex if not impossible!

5.3 SURVEY OF ROTATION TECHNIQUES

In general there are two distinct categories of factor rotation--orthogonal and oblique--which differ widely both in concept and interpretation. The idea of strictly uncorrelated factors in the orthogonal structure, whether simple structure or not, has contributed significantly to the extensive usage of the orthogonal solution in a final analysis. Simply summing the squares of all the factor loadings for any given variable yields its common factor variance, thus, the importance of an individual loading is easily determined. This is not at all the case in oblique factor structures where nonzero correlations among factors necessitate rather tedious matrix manipulations which heavily tax the skills and patience of the user. A simpler method to determine factor significance is not yet known but the problem is considered later on in this section.

It is indeed unfortunate that interpertation ease has dictated the unquestioned popularity of the orthogonal methods since the shortcomings of a linear model are confounded by a further restriction of uncorrelated factors. A more realistic model (naturally there are many problems which

fall into the "straightforward" class whereby all of the common factor variance can be accounted for by the first few factors and furthermore can be interpreted as definite orthogonal physical factors) in the physical world is, of course, the oblique factor structure if the intent of the analyses is one of discovering physical entities. In data reduction problems the orthogonal patterns are quite acceptable.

Helpful in the selection and comparison of simple structure rotation techniques is Table 6 extracted in part from Harman (Reference 2, p. 310) where short expressions for Quartimax and Varimax orthogonal rotation techniques and Oblimax, Quartimin, Covarimin, Oblimin, and Kaiser-Dickman oblique rotation techniques are given. The following notation is adopted for the table:

(a_{jp}) = initial factor matrix,

(b_{jp}) = final factor matrix,

(v_{jp}) = final factor structure matrix.

It should be noted that major differences in these techniques occur both in concept of a "best" simple structure and in computation procedures. The orthogonal rotation problem is pretty well resolved by Varimax, Quartimax at best being a good estimate. The oblique techniques require enormous computation efforts and generally result in "not quite" solutions which call upon Cattell's Maxplane, or Rotoplot, for polishing.

Table 6

Rotation Methods and Their Criteria

Method	Orthogonal Case	Oblique Case
Quartimax	Equivalent expressions:	
	$Q = \sum_{j=1}^{n} \sum_{p=1}^{m} b_{jp}^4 = \max.$	
	$M = \frac{1}{mn}\left[\sum_{j=1}^{n} \sum_{p=1}^{m} b_{jp}^4 - \sum_{j=1}^{n} \sum_{p=1}^{m} b_{jp}^2\right] = \max.$	
	$N = \sum_{p<q=1}^{m} \sum_{j=1}^{n} b_{jp}^2 b_{jq}^2 = \min.$	
	$K = \sum_{j=1}^{n} \sum_{p=1}^{m} b_{jp}^4 \Big/ \left(\sum_{j=1}^{n} \sum_{p=1}^{m} b_{jp}^2\right)^2 = \max.$	
Varimax	$V = n \sum_{p=1}^{m} \sum_{j=1}^{n} (b_{jp}/h_j)^4 - \sum_{p=1}^{m} \left(\sum_{j=1}^{n} b_{jp}^2 / h_j^2\right)^2 = \max.$	

Table 6 Cont.

Method	Orthogonal Case	Oblique Case
Oblimax		$K = \sum_{j=1}^{u} \sum_{p=1}^{m} v_{jp}^4 \Big/ \left(\sum_{j=1}^{n} \sum_{p=1}^{m} v_{jp}^2 \right)^2 = \max.$
Quartimin		$N = \sum_{p<q=1}^{m} \sum_{j=1}^{n} v_{jp}^2 v_{jq}^2 = \min$
Covarimin (oblique varimax)		$C = \sum_{p<q=1}^{m} \left[n \sum_{j=1}^{n} (v_{jp}^2/h_j^2)(v_{jq}^2/h_j^2) - \sum_{j=1}^{n} v_{jp}^2/h_j^2 \sum_{j=1}^{n} v_{jq}^2/h_j^2 \right] = \min.$
Oblimin		$B = \sum_{p<q=1}^{m} \left[n \sum_{j=1}^{n} (v_{jp}^2/h_j^2)(v_{jq}^2/h_j^2) - \gamma \sum_{j=1}^{n} v_{jp}^2/h_j^2 \sum_{j=1}^{n} v_{jq}^2/h_j^2 \right] = \min,$

Table 6 Cont.

Method	Orthogonal Case	Oblique Case
Kaiser-Dickman		where $\gamma=\beta/(\alpha+\beta)$ and α and β are weights. $$D=\sum_{p<q=1}^{m}\left[\frac{\sum_{j=1}^{n}(v_{jp}^2/h_j^2)(v_{jq}^2/h_j^2)}{\left(\sum_{j=1}^{n}v_{jp}^2/h_j^2\right)\left(\sum_{j=1}^{n}v_{jq}^2/h_j^2\right)}\right]=\min.$$

5.4 INTERPRETATION OF OBLIQUE FACTORS

A. Introduction

The factor analysis model involves the simultaneous linear description of n variables by m common factors and n unique factors.

$$\begin{aligned} Z_1 &= a_{11}F_1 + a_{12}F_2 + \cdots + a_{1m}F_m + a_1U_1 \\ &\vdots \\ Z_j &= a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + a_jU_j \\ &\vdots \\ Z_n &= a_{n1}F_1 + a_{n2}F_2 + \cdots + a_{nm}F_m + a_nU_n \end{aligned} \qquad (1)$$

The factors are, of course, hypothetical and their description is usually given by a pattern matrix $A = (a_{jp})$ of common factor coefficients, and a structure matrix $S = (s_{jp})$, the set of correlations between each variable and factor.

The invariant part of a factor analysis solution is the subspace of common factors, common-factor space, defined by the set of standardized column vectors $F_1, \ldots, F_m$. The n-space of standardized variables Z_j lies hopefully close to the m space of common factors and each variable is projected onto common-factor space by its unique factor, a_jU_j. Selecting a particular solution for the factor analysis model corresponds to selecting a set of basis vectors $\{F_1, \ldots, F_m\}$ to describe the invariant common-factor space.

The projection of variable Z_j on common-factor space is $\tilde{Z}_j$, the prediction of Z_j from the common factors alone.

$$\tilde{Z}_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m , \qquad (2)$$

so that

$$Z_j = \tilde{Z}_j + a_jU_j . \qquad (3)$$

Thus the variance of Z_j is

$$(Z_j|Z_j) = 1 = (\tilde{Z}_j|\tilde{Z}_j) + 2a_j(\tilde{Z}_j|U_j) + a_j^2(U_j|U_j) \qquad (4)$$

The variance of Z_j is called its communality h_j^2. Because the unique factors are orthogonal to all common factors, $(\tilde{Z}_j|U_j) = 0$ and since the variance of the unique factors is one, Equation 4 becomes

$$\text{var}\,(Z_j) = 1 = h_j^2 + a_j^2 \qquad (5)$$

The communality h_j^2 is the variance explained by the common factors. Both h_j and a_j are fixed for any factor analysis solution, hence for the entire set of particular solutions generated by rotating the common factors to different bases for common-factor space.

B. Problems with Oblique Factors

In order to understand intuitively the dimensions of common-factor space or to identify factors it seems likely that an oblique set of factors is preferable. In addition, a factor which has been placed close to a group of real, observed variables would seem more likely observable itself.

However, there are serious problems involved in the interpretation of the output of oblique rotations which have discouraged many workers from leaving orthogonality. The pattern and structure matrices are not identical and they are both tricky. For some examples let us consider two-factor space: $\tilde{Z}_j = a_{j1}F_1 + a_{j2}F_2$. A variable may be uncorrelated with a factor F_1 and yet have a high loading a_{j1} on it or it might have a large positive structure value s_{j1} and yet a negative loading a_{j1}. (See Figure 6.)

The basic difficulty in interpreting the structure matrix, and part of the reason for these seeming discrepancies between structure and pattern, is that the variable-factor correlations are affected in the oblique case by the correlations among factors. This will be explained in more detail later. Moreover, two variables may be correlated with one

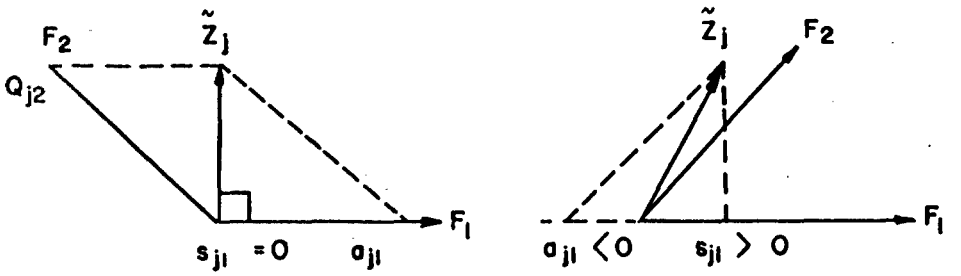


Figure 6. Relation between Factors and Variables

factor as high as .707 and yet be totally uncorrelated themselves, making it difficult to pick out groups of variables and fit them to factors by examining structure alone.

The problem with oblique factor patterns is that the sum of the loadings squared for one variable

$$\sum_{p=1}^{m} a_{jp}^2$$

must no longer equal the communality or even be less than one, as in the orthogonal case. Although rotation tends to purge middle-sized loadings, it may result in loadings greater than one, or in several large loadings which indicate not so much linear determination as they do that the factors are highly uncorrelated (or correlated) as in Figure 6.

We may state the problem: given the linear representation of $\tilde{z}_j$ as in Equation 2

$$\tilde{z}_j = a_{j1}F_1 + a_{j2}F_2 + \ldots + a_{jm}F_m$$

how "important" is each factor in the determination of all the values of the vector $\tilde{Z}_j$.

The contribution to the variance of $\tilde{Z}_j$ is the measure usually used. The variance of $\tilde{Z}_j$ is

$$(\tilde{Z}_j|\tilde{Z}_j) = h_j^2 = \sum_{p=1}^{m}\sum_{q=1}^{m} a_{jp}a_{jq}(F_p|F_q) = \sum_{p=1}^{m}\sum_{q=1}^{m} a_{jp}a_{jq}r_{F_pF_q} \quad . \qquad (6)$$

When the factors are orthogonal, it is

$$r_{F_pF_q} = (F_p|F_q) = \begin{cases} 0, & p\neq q \\ 1, & p=q \end{cases}$$

so that

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 \quad . \qquad (7)$$

Thus the contribution of each <u>orthogonal</u> factor to the variance of $\tilde{Z}_j$ is the square of its loading. This clean resolution of variance explains why orthogonal factors may be easily interpreted; their relative importance in determining all the values for a variable can be separately evaluated.

However, for oblique factors the terms containing $r_{F_pF_q}$, $p\neq q$, do not drop out and we have

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \cdots + a_{jm}^2 + 2\left(a_{j1}a_{j2}r_{F_1F_2} + \cdots + a_{j1}a_{jm}r_{F_1F_m}\right) + 2\left(a_{j2}a_{j3}r_{F_2F_3} + \cdots + a_{j2}a_{jm}r_{F_2F_m}\right) + \cdots + 2\left(a_{j,m-1}a_{jm}r_{F_{m-1}F_m}\right) \qquad (8)$$

The terms a_{jp}^2 are called direct contributions. The mixed terms may be named <u>two-factor interactions</u> and they cause the problems. These interactions are not variance "contributions" because, for one thing, they may be negative. They may be looked upon best as corrections

applied to the direct contributions a_{jp}^2 due to the tendancy of one factor to vary with another.

One method of separating the total variance into components for each factor is to simply divide the interaction into halves and assign a half to each factor. This results in:

$$h_j^2 = \sum_{p=1}^{m} a_{j1} a_{jp} r_{F_1 F_p} + \cdots + \sum_{p=1}^{m} a_{jm} a_{jp} r_{F_m F_p}$$

$$h_j^2 = a_{j1} \sum_{p=1}^{m} a_{jp} r_{F_1 F_p} + \cdots + a_{jm} \sum_{p=1}^{m} a_{jp} r_{F_m F_p}$$

$$h_j^2 = a_{j1} s_{j1} + \cdots + a_{jm} s_{jm} \quad . \qquad (9)$$

The terms of Equation 9 might be said to approximate the contribution of each factor to the variance of $\tilde{Z}_j$. In the orthogonal case it reduces to Equation 7 and it often gives an enticingly clear picture. But it also results in negative values whenever a_{jp} and s_{jp} are of opposite sign. This is one indication that the interactions are simply that--interactions--and cannot be resolved into shares.

Any procedure such as halving the interaction which gave us a matrix of contributions to variance could generate an orthogonal pattern simply by taking the square root of each element. Therefore, because each set of factors has only one pattern, A, no oblique factors may be so resolved. When the factors are correlated, the analysis of variance model (separating the sums of squares) can no longer be used but a new model must be formulated. (Reference 61, p. 464, p. 634).

C. Factor Analysis and Regression Analysis

To facilitate the development of a statement concerning the contribution of oblique factors to variance, let us show that the factor pattern equations are a set of classical regression equations of the variables on m common factors.

The factor analysis description of a variable Z_j

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + \cdots + a_{jm}F_m + a_jU_j \qquad (10)$$

is simply a linear equation of the variable Z_j in terms of $m+1$ others. It makes no difference mathematically that the one is observed and the others hypothetical. The value of Z_j predicted by the common factors is $\tilde{Z}_j$.

$$\tilde{Z}_j = a_{j1}F_1 + \cdots + a_{jm}F_m \qquad (11)$$

Harman (Reference 2, p. 18) proves that Equation 10 is a regression equation but we may just as easily prove the more pertinent theorem that Equation 11 is a regression equation.

The sum of squares of residuals for Equation 11 (over the N values of the vector) is

$$\sum_{i=1}^{n} (Z_{ji} - \tilde{Z}_{ji})^2 = \left|(Z_j - \tilde{Z}_j) \mid (Z_j - \tilde{Z}_j)\right| = a_j^2$$

Since $a_j^2 = 1 - h_j^2$, $j = 1, \cdots, n$, are unique for any factor analysis solution they may be regarded as at a minimum for a set of factors (this assumption also defines the factors as least squares estimates). Hence Equation 11 may be regarded as a least squares solution and a regression equation with a standard error of estimate a_j.

A regression equation is usually represented:

$$Y = b_1x_1 + b_2x_2 + \cdots + b_mx_m + \varepsilon$$

or simply in vector notation

$$\tilde{Y} = XB \qquad (12)$$

where $\tilde{Y}$ is a least squares estimate to Y; $\tilde{Y}$, Y, X are score vectors: B is a coefficient vector. We may easily imagine Y and B extended to matrices of column vectors. The factor analysis model uses row vectors for scores and linear coefficients. Assuming that all variables are standardized we may let

$$Z' = Y$$
$$F' = X$$
$$A' = B$$

and

$$\tilde{Z} = AF$$
$$\tilde{Z}' = F'A' \text{ is equivalent to } \tilde{Y} = XB \ .$$

Thus the factor pattern is a set of regression coefficients.

The regression analysis solution for B is

$$X'\tilde{Y} = X'XB$$
$$B = (X'X)^{-1}X'Y \qquad (13)$$

The C matrix is usually defined

$$C = (X'X)^{-1}$$

and then

$$B = C\,X'Y \ .$$

In factor analysis we are given

$$(X'X) = FF' = \Phi \qquad (14)$$

so that

$$C = \phi^{-1}$$

and any uses for C may be referred to ϕ^{-1}, the inverse of the matrix of factor correlations.

The <u>multiple correlation</u> of Y_j on $X_1, \ldots, X_m$ is defined as

$$r_{\tilde{Y}_j Y_j}$$

(Section 2.7). Thus, letting q stand for all X_p, p = 1, ..., m, and using standard notation for multiple correlation:

$$R_{Y_j \cdot X_1 \ldots X_m} = R_{Y_j \cdot q} = \frac{(Y_j|\tilde{Y}_j)}{\sqrt{(Y_j|Y_j)(\tilde{Y}_j|\tilde{Y}_j)}}$$

But

$$Y_j = \tilde{Y}_j + a_j U_j \ , \ (\tilde{Y}_j|U_j) = 0 \ , \text{ and } (Y_j|Y_j) = 1$$

so that

$$R_{Y_j \cdot q} = \frac{(\tilde{Y}_j|\tilde{Y}_j)}{\sqrt{(\tilde{Y}_j|\tilde{Y}_j)}} = \sqrt{(\tilde{Y}_j|\tilde{Y}_j)} = h_j \qquad (15)$$

Hence the squared multiple correlation of a variable on the m common factors is its communality, or its explained variance.

D. A Proposed Measure for the Importance of Oblique Factors

Methods usually associated with regression analysis enable us to make a statement regarding the contribution of correlated factors to the variance of their dependent variable. It concerns not the direct contribution which we have shown to be meaningless in the oblique case, but the amount of explained variance which a factor adds after all others have been taken into account. Although this measure is probably as much as can be said about the separate effect of a factor, it is a natural and useful statement.

As stated before, the variance due to oblique factors cannot be simply divided among them, due to the two-factor interactions--the tendency of factors to vary together. Hence we might search for a way to examine the relationship between a factor and a variable with the other factors held constant. We tend to assume that if a variable and a factor are correlated that the factor is (mathematically) affecting the variable; but with correlated factors the observed correlation between variable and factor may be spurious--the results of limitations placed on their

correlation by being both tied to a second factor. Ezekiel (Reference 23, p. 195) states the other possibility: "It is evident that a mere surface examination of a set of data cannot reveal which independent factors are important and which are unimportant. A factor which shows no correlation with the dependent variable may yet show significant correlation after the relation to other variables has been allowed for."

Consider three correlated variables. If the correlation of two variables were measured for groups of fixed values of the third variable and a weighted average formed, the correlation would probably be different. We called such a measure partial correlations (Section 2.7) and it may be written in terms of simple correlations.

$$r_{12.3} = \frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}} \qquad (16)$$

This value is the correlation of one and two with the disturbing effect of three removed. (A proof of Equation 16 may be found in Reference 62, p. 479) The relationship between partial correlation and multiple correlation (explained variance) is given as follows (Reference 63, p. 344) for the regression of Y on 4 factors

$$1-R^2_{Y.1234} = (1-r^2_{Y1})(1-r^2_{Y2.1})(1-r^2_{Y3.12})(1-r^2_{Y4.123}) \qquad (17)$$

This expression may be extended to the regression of Y on X_1 to X_m by multiplying the right hand side of Equation 17 by appropriate terms in the series. Thus if we let q represent all factors but those to the left of the dot and in parentheses beside q

$$1-R^2_{Y.q} = \left|1-R^2_{Y.q\ (X_p)}\right| \left| 1-r^2_{YX_p.q}\right|$$

The order of X's does not matter; X_p may be any of the factors $X_1, \cdots, X_m$. Solving for the partial correlation squared

$$r^2_{YX_p.8} = \frac{\left|1-R^2_{Y.q(X_p)}\right| - \left|1-R^2_{Y.q}\right|}{\left|1-R^2_{Y.q(X_p)}\right|} \qquad (18)$$

For four factors the partial correlation squared between Y and X_2 becomes

$$r^2_{Y2.134} = \frac{(1-R^2_{Y.134})-(1-R^2_{Y.1234})}{(1 - R^2_{Y.134})} . \tag{19}$$

Simplifying the numerator of Equation 19 yields $R^2_{Y\cdot1234} - R^2_{Y\cdot134}$. Recalling that multiple correlation squared equals communality and generalizing to m factors, the numerator of Equation 19 is seen to be the difference between the explained variance of Z_j as a regression on all m common factors and its variance as a regression on all the common factors but one. Let this difference for the omission of factor F_p (X_p is equivalent regression language) be denoted by v^2_{jp}. This is the proposed measure. It has shown up while examining the relationship between a variable and a factor with the disturbing effect of other factors, due to co-variance, removed, as it is implicitly with independent factors.

We may define the <u>unique contribution to variance</u> v^2_{jp} of factor F_p for variable Z_j as the <u>additional</u> variance explained by factor p after all the variance of Z_j explainable by the other factors $F_k, k \neq p$ has been taken into account. More formally

$$v^2_{jp} = R^2_{Y_j.F_1,\ldots,F_m} - R^2_{Y_j.F_1,\ldots,F_{p-1}(F_p)F_{p+1},\ldots,F_m}$$

$$= R^2_{Y_j.q} - R^2_{Y_j.q(F_p)} \tag{20}$$

and we state a theorem which is proved at the end of this section. When v^2_{jp} is the unique contribution to variance as defined above

$$v^2_{jp} = \frac{b^2_{jp}}{c_{pp}} \tag{21}$$

or in factor analysis notation

$$v_{jp}^2 = \frac{a_{jp}^2}{\phi_{pp}^{-1}} \qquad (22)$$

Due to the interactions this expression is about as much as can be said about the separate "importance" of the oblique factors to the explained variance. It represents the part of the total variance which must be explained by that factor or be lost--a natural and meaningful measure.

Furthermore, the coefficient v_{jp}^2 is a generalized measure for all sets of factors, orthogonal ones being a special case which happens to sum to the total explained variance. This phenomenon exists because the interactions are zero and thus a factor's contribution cannot be particularly picked up by another factor. Notice that then $\phi^{-1} = I$ and Equation 22 reduces to

$$v_{jp}^2 = a_{jp}^2$$

For oblique factors

$$\sum_{p=1}^{m} v_{jp}^2 < h_j^2$$

which indicates again that part of the explained variance is not "unique" to any one factor.

Using Equations 15, 19, and 20 the partial correlation between Y_j and F_p becomes

$$r_{Y_j F_p \cdot q} = \sqrt{\frac{v_{jp}^2}{1-h_j^2 + v_{jp}^2}}$$

It is a measure of the correlation with other factors held constant, and it might be used in a "corrected" structure matrix to help name the factors. The square of this term is seen to be the unique contribution to variance divided by the variance of Z_j with F_p removed, or simply how much (a ratio) of the otherwise unexplained variance it explains. It is used by several authors as a measure of the importance of factor F_p. However, v_{jp}^2 seems to be a stronger measure because it is desirable in factor

analysis to keep our "importance coefficient" in terms of variance and absolute for inter-variable comparisons. Thus for evaluating the importance of a factor v_{jp}^2 is preferable.

It is important to remember that the contribution to variance of an oblique factor is not a unique value but a range of possible values. We may think of v_{jp}^2 as a sort of lower bound to this range. Perhaps it would be worthwhile to also set an upper bound, or to examine the consequences if two or three factors were removed at a time. Further investigation of the problem is needed. This section offers v_{jp}^2 as an easily computable measure of the unique contribution to variance. Perhaps it and other measures to come can put oblique rotations on the road to engineering practice.

Proof of Equation 22: The elements of any j^{th} row (or column) of $(X'X)^{-1} = C=\phi^{-1}$ divided by the negative of diagonal element c_{jj} give the regression equation of X_j in terms of the other X's.

$$X_j = \frac{c_{j1}}{c_{jj}} X_1 - \cdots - \frac{c_{j,j-1}}{c_{jj}} X_{j-1} - \frac{c_{j,j+1}}{c_{jj}} X_{j+1} - \cdots - \frac{c_{jm}}{c_{jj}} X_m \qquad (23)$$

Let us prove Equation 23 in more useful terms, using the column vector Y instead of X_j and letting X be a set of column vectors $X_1,\ldots,X_m$, from the partitioned matrix $[Y|X]$
Then let

$$F = [Y|X]'[Y|X] = \left[\frac{Y'}{X'}\right][Y|X] = \left[\frac{Y'Y|Y'X}{X'Y|X'X}\right]$$

(F is simply the matrix of correlations for the columns of $[Y|X]$ as ϕ is for X).
Let the partitioned inverse of F be

$$\left[\frac{e \mid D}{D' \mid E}\right] = F^{-1} = ([Y|X]'[Y|X])^{-1}.$$

Then

$$\left[\begin{array}{c|c} Y'Y & Y'X \\ \hline X'Y & X'X \end{array}\right] \left[\begin{array}{c|c} c & D \\ \hline D' & E \end{array}\right] = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & I \end{array}\right] \qquad (24a)$$

and one of the four resulting equations is

$$X'Ye + X'XD' = 0 \qquad (24b)$$

therefore

$$-\frac{1}{c} D' = (X'X)^{-1} X'Y = B \; . \qquad (24b)$$

But the expression in the middle of Equation 24b is the least squares solution for the regression coefficients B in Equation 13 and D is a column of F^{-1}. Hence Equation 23 is true. We may also show from Equations 24a and 24b that

$$Y'Ye + Y'XD' = I \; .$$
$$Y'Ye - Y'XBe = I \; .$$

Hence

$$e = (Y'Y - Y'XB)^{-1} \; .$$

For standardized variables $Y'Y$ is one; and Y' is a row vector of values while XB is a column vector of least squares estimates of those values.

Thus

$$Y'XB = Y'\tilde{Y} = (Y|\tilde{Y}) = R^2_{Y.q}$$

and

$$e = \frac{1}{1 - R^2_{Y.q}} \qquad (25)$$

This is the well known fact that the diagonal elements of R^{-1} contain the multiple correlation of one variable on all the others.

Now we may examine the effect on a least squares regression equation of eliminating a factor by studying the effect of eliminating a row and column (before inverting) on the inverse of a correlation matrix.

Specifically, let $G^{-1} = (g_{ij})$ be the $m \times m$ matrix of the inverse of G where G is formed from $F = (f_{ij})$ as before by eliminating factor X_u.

Brownlee (Reference 62, p.489) gives us the formula

$$g_{ij} = f_{ij} - \frac{f_{iu}f_{ju}}{f_{uu}} \quad . \tag{26}$$

In particular for the diagonal elements of G^{-1} corresponding to vector Y

$$g_{yy} = f_{yy} - \frac{f_{yu}^2}{f_{uu}} \quad . \tag{27}$$

The matrix C as defined previously may be thought of as the inverse of the matrix of F after removing the variable Y.

Thus Equation 26 becomes

$$c_{ij} = f_{ij} - \frac{f_{iy}f_{jy}}{f_{yy}} \quad .$$

Letting $i=j=u$ and solving for f_{uu} in the above equation

$$f_{uu} = c_{uu} + \frac{f_{uy}^2}{f_{yy}} \quad . \tag{28}$$

Equation 27 may be written

$$g_{yy} = \frac{f_{yy}f_{uu} - f_{yu}^2}{f_{uu}} .$$

Substituting Equation 28 for f_{uu} and noting that $f_{yu} = f_{uy}$

$$g_{yy} = \frac{f_{yy}^2 c_{uu}}{c_{uu}f_{yy} + f_{yu}^2} . \qquad (29)$$

According to Equation 23

$$\frac{f_{yu}}{f_{yy}} = -b_{yu}$$

where b_{yu} is the regressive coefficient of Y on X_u .

But f_{yy} and g_{yy} are diagonal elements equivalent to e in Equation 25. Hence

$$\frac{1}{g_{yy}} = 1 - R^2_{Y.q(X_u)} .$$

and

$$\frac{1}{f_{yy}} = 1 - R^2_{Y.q} ,$$

so that Equation 29 becomes

$$\frac{1}{g_{yy}} = \frac{1}{f_{yy}} + \frac{f_{yu}^2}{f_{yy}^2 c_{uu}} = 1 - R^2_{Y.q(X_u)} = 1 - R^2_{Y.q} + \frac{b_{yu}^2}{c_{uu}}$$

or in factor analysis notation for a particular Z_j after removing F_p

$$v_{yp}^2 = R_{Y_j.q}^2 - R_{Y_j.q(F_p)}^2 = \frac{b_{jp}}{c_{pp}} = \frac{a_{jp}^2}{\phi_{pp}^{-1}}$$

which is Equation 22.

For m=3 and p=1 Equation 22 may be written

$$R_{Y.123}^2 - R_{Y.23}^2 = \frac{b_{y1}^2}{c_{11}}$$

and this equation may be found in Reference 63, p. 339.

Section VI

UNIQUENESS OF FACTOR ANALYSIS

6.1 INTRODUCTION

Section VI deals with the problem of uniqueness in factor analysis. The concept of uniqueness is described in subsection 6.2. Uniqueness, on the other hand, is closely related to the much more practical problem of how large a sample one has to have for doing a factor analysis. So 6.3 shows this relationship and then establishes a means how to solve the two problems, which are actually the one problem of uniqueness in factor analysis.

6.2 THE ISSUE OF UNIQUENESS

The issue of uniqueness can be described as follows. Two independent teams are told to collect data and perform a factor analysis of a certain subject matter area. The issues are described in the same way to each team. Data collection and analysis is performed independently by each team, independent decisions are made about factoring, and separate final reports are drawn up. The issue of uniqueness is this: will the reports be "basically" the same?

Of course, the issue has been transformed into one centering on the meaning of "basically." If the picture is redrawn slightly the issues will be clearer. Suppose, to make it more specific, that the study is the psychological one mentioned above, focused on one large school, and using examination results of students to uncover mental factors. As we now impose more conditions on the picture, the reports of the teams ought to grow more and more similar. First we require that neither team make longitudinal studies, then we require that there should be no separate analysis for males or females, nor for school grades. Next we require that neither team invent and administer to its own test on, say manual dexterity or reading speed. Finally, we require that each team use the same squared multiple correlation for communalities, and varimax rotation.

It should be clear that continuing to standardize the teams will lead us to the point that any discrepancies between the final reports must be

due to sampling--one team selected one group of students, and the other team selected another group. There may have been an overlap, but still the final reports are different.

Let us now make one further change in the foregoing picture. Suppose that one team has studied boys only, and the second has studied girls only, and we wish to know whether the differences in their final reports are due to sampling differences, or to sex differences. Here we are at the crux of the issue of uniqueness. If the two teams were measuring some simple statistic, like the classroom grade or height, the issue could be simply resolved by the appropriate F-test or t-test, but in factor analysis we are dealing with a highly complex set of interrelated statistics.

6.3 SAMPLING CONSIDERATIONS

Although it may not appear so at first sight, the issue of uniqueness is also very closely related to the much more practical problem of how large a sample one ought to work with in a factor analysis. Of course if observations are cheap, there is no problem, and the issue is resolved by considering the clerical facilities available for copying or punching numbers. Bad data can be freely edited out, and there is only one question facing the investigator: is the data really representative of the population of response about which I wish to make inferences? More specifically, the issue can be rephrased as follows: when an outlier is thrown out because it is unrepresentative, can I be sure that I am drawing inferences about a population which is also free of "unrepresentative" observations? If the answer is no, then the investigator should not throw out such data.

Usually observations are expensive to collect, and one cannot simply choose 1000 observations because it is a round number. The investigator must select one sample of 200, say, and remind himself that another investigator in doing a similar factor analysis might have selected a different sample of 200 to work with. The conclusions of one investigator should not contradict those of the other, no matter whether the second investigator is real or imaginary. So here again we are at the same issue as that posed previously as the issue of uniqueness.

This time phrased non-technically, we call it the issue of sampling variability. Much of the purely mathematical statistical issue has been resolved, but the conclusions have not yet been formulated in terms of rules of thumb which the non-professional can use.

To develop such rules, a simulation program has been written. Basically, it set several imaginary investigators to work on the same data, as described above. Differences between the results of these investigators are then examined, and in this way we can discover how tentatively one investigator must describe his results in order not to contradict or be contradicted by an imaginary colleague.

Since we shall assume that the data available to investigators are normally distributed, the starting point of any such simulation will be a need to generate multivariate normal deviates in the computer. This issue does not seem to have been dealt with directly in the literature, but can be solved in the following manner. Let us assume that random normal deviates are available as needed. These can be generated directly through any of the methods now available, or generated indirectly through a random (rectangular) number generator plus a "normit" routine which provides a normal deviate corresponding to any desired probability level. The probability level of course will be obtained from the random number generator. With such random deviates $\underline{x}$ freely available, drawn from a standardized population with mean zero and variance unity, we desire to generate a multivariate normal vector variable $\underline{y}$ which shall be standardized to zero mean and unit variance, but shall have any prescribed covariance, i.e., correlation, structure R. The positive definite correlation matrix R, of size $n \times n$, will then describe the population which we are factor analysing. If R is the unit matrix, then $\underline{y} = \underline{x}$ will serve as the generated variable, but in general it will be necessary to discover the non-symmetric matrix A, of size $n \times n$, which has the property that

$\underline{y} = A\underline{x}$ has correlation matrix R.

The covariance of the vector variable $\underline{y}$ is given by

$$E(\underline{y}'y) = E(\underline{x}' (A'A)\underline{x}).$$

Using the fact that $\underline{x}$ is standardized, we can expand the condition that $\underline{y}$ have matrix $R = (r_{ij})$ by expressing it in terms of conditions on $A = (a_{ij})$.

There are enough degrees of freedom that we may immediately impose the condition that $\underline{y}$ be standardized. In this case the covariance and correlation of $\underline{y}$ are identical and the simple condition that A must satisfy that

$$A'A = R.$$

Expanding this, it is a system of equations

$$\sum_k a_{ki} a_{kj} = r_{ij}$$

with diagonal elements, specifically,

$$\sum a_{ki}^2 = r_{ii} = 1$$

and $r_{ij} = r_{ji}$

following trivally from the above expansion.

A simple example will illustrate the situation here. If $n = 2$, we have two independent standard normal deviates x_1 and x_2, and wish to manufacture two other variates y_1 and y_2 with the property that they have a desired correlation r with each other. It can easily be verified that if y_1 and y_2 are defined as

$$y_1 = x_1$$

$$y_2 = rx_1 + \sqrt{1-r^2} \cdot x_2$$

they will have the desired property. The matrix A thus defined is easily constructed for $n = 2$. For $n = 3$ it will be seen that the

algebra becomes complicated. There will be three specified correlations r_{12}, r_{13} and r_{23}. Building up the desired variates as before, we have

$$y_1 = x_1$$

$$y_2 = r_{12}x_1 + \sqrt{1 - r_{12}^2}\, x_2$$

$$y_3 = r_{13}x_1 + \frac{r_{23} - r_{12}r_{13}}{\sqrt{1 - r_{12}^2}}\, x_2 + \sqrt{1 - r_{13}^2 - a^2}\cdot x_3$$

where a is the coefficient of x_2 in y_3. Although the algebra rapidly becomes impossibly complex, the process is very straightforward and can easily be built into an algorithm for use in a computer. Specifically, if i exceeds j, a will be developed in such a way as to produce the desired correlation r_{ij}, and if i equals j, the coefficient will be developed so as to ensure unit variance of the corresponding y.

The above transformation, or one similar to it, has been used in other connections by various authors, but ordinarily for the opposite purpose, namely to provide uncorrelated variates from correlated ones.

It is important to note that the transformation will produce conservative results. That is, the correlation matrix will be treated as if it were a population matrix, even though it is only a sample matrix. Thus the later sample correlation matrices which are developed by the algorithm will be more like the original sample matrix than they really "ought" to be. The only alternative would involve building a model based on "fiducial" distributions of population parameters, and strong exception would be taken to this procedure by many investigators. The results coming from the program are striking enough that the conservatism is not objectionable.

A computer program, within the UNIVAC 1105, has been developed incorporating the above algorithm. Basically it contains the following prodedure.

1. Reads the input correlation matrix R , the sample size N to be used, the matrix size r , and the number of iterations to be made of the program (see below).

2. Generates N vector variables $\underline{y}$ which are pseudo-random samples from a normal population with correlation matrix R .

3. Forms the correlation matrix of these variables.

4. Calculates the squared multiple correlation estimates of the r communalities, perform the principal-factor solution to the factor analysis.

5. Prints out the sample correlation matrix and associated communality estimates, and the characteristic roots and scaled vectors of the solution. Saves the answers in computer binary format to be used below and in the varimax rotation program.

6. Repeats steps 2 to 5 above the number of times requested in iteration parameter in step 1 above.

7. Calculates the averages and variances of all the eigenvalues and eigenvectors and prints these out.

8. Returns to step 1 above unless directed to terminate the program.

The purpose of the program was two-fold. First, to see how the stability of estimates increases as sample size N increases, and secondly to see how this same stability is influenced by the correlation structure. Most of the runs were performed with independent population data, so that roots and vectors were calculated from data with the identity correlation matrix. The number of iterations under various parameter combinations is given in the following table. A non-orthogonal design was used because of the machine time and costs associated with large variate sizes, and was close to optimum when these costs are considered as part of the design.

Table 7

Number of Computer Runs of Factor Analysis
on Independent Data

No. of variates	Sample Size 100	200	300	Total
30		3		3
20	11			11
10	6	6	6	18
Total	17	9	6	32

In addition, 11 iterations were made on the classical 24 psychological variate test data from the Spearman-Holzinger Unitary Trait Study, used by Harman (Reference 2) and others. The same N of 145 was used as in the original study, and of course the number of variates, n, was taken as 24.

Some general conclusions of practical relevance are as follows.

1. When we are sampling from independent data, the use of squared multiple correlation (SMC) communalities tends to create "errors of the first kind." That is, it leads to production of one or even two roots which are relatively larger than all the others. For instance, in one of the three iterations for $N = 300$, $n = 30$, the two largest roots were .95 and .91, followed by much smaller roots .37, .25, .23, etc. To describe the situation intuitively, what happens is that if the sampling from independent data appears by accident to produce something that looks significant, the SMC procedures jump on it and try to make it look good.

2. When SMC communalities are used, common practice is to assume that there will be one insignificant positive root for every negative root. From the previous paragraph, it can be further suggested that one and possibly two further small positive roots can be assumed insignificant because of the SMC bias mentioned above. Of course, unless the roots are much in excess of unity there can be no significance imputed to them in any case. Insignificant roots tend to be largest when N is small

and n is large, as one might expect. When N = 100, n = 20, for instance, half of the largest roots exceeded unity, but none was larger than 1.18.

3. The sampling variability of the largest root in independent data is surprisingly independent of both N and n. The variance is approximately .01. It is much larger if the root is significant, but the coefficient of variation, i.e., the standard deviation relative to the mean remains relatively stable at about 10 per cent. This can serve as the rule of thumb for largest eigenvalues.

4. The sampling variability of the eigenvectors corresponding to the largest root in independent data depends on both N and n. The variance of the eigenvectors decreases approximately as the inverse of the square root of N. Thus it is relatively insensitive to changes in N. To illustrate, if N = 100, n = 10, the variance is .048 and the standard deviation of course is .22. If we take four times as many observations, the variance is reduced by one-half, and the corresponding standard deviation is .155, not a great improvement on .22 considering the quadrupling of data involved.

This sampling variance also diminishes as n increases. However, the relationship here is much more complex, and the experimental design used in collecting the data did not permit high clarification on this point. As a tentative approximation, it appears that the sampling variance diminishes as the inverse of n.

The foregoing conclusions seem to be at variance with those of Harman (Reference 2, Appendix, Table B), but comparison is not possible since his results do not apply to independent data. From practical experience, it seems desirable to make the pessimistic assumption that the numerical information one has collected does not look encouraging and that the investigator would be happy to find any significant pattern at all in it.

The rule of thumb suggested from the foregoing is that the variance of the eigenvectors corresponding to largest roots in independent data is $5/n\sqrt{N}$.

5. When the data contain significant material, the sample eigenvectors do have a population value to gravitate towards, and hence the sampling variability of the coefficients diminishes. In the Guilford trait data, the

variance of the eigenvectors corresponding to the largest root was .004. For smaller roots, where sampling error only is being measured, the variance increases to where it corresponds to that of independent data, as one would expect. This variance of .004 is only one-tenth what one would expect if the same parameters had operated on independent data.

Unfortunately, no dependable rule of thumb can be inferred which would apply to all data. It will depend on how strong the population eigenvector is to which the sample is tending. As a very crude first approximation, one might measure this strength by means of the largest eigenroot V, and hence adopt $5/Vn\sqrt{N}$ as a rule of thumb for the variance.

One further important possibility opens to the investigator because of the relative unimportance of the size of N. If the investigator has say 400 observations, he can do one analysis on all the data, and then divide the data in half at random and do separate analyses on each half as well. Because the data from 200 observations will be nearly as "good" as that from 400, it follows that any factor that seems to appear in the analysis of the 400 observations is dependable only if it can also be discerned in the analysis of <u>each</u> of the two halves of the data.

6. The most unexpected result of this investigation is that with the use of SMC communalities on the Unitary Trait data, it can be statistically established through the sampling scheme used here that there is only one significant factor in the data, rather than four (e.g., Harman, Reference 2, Table 9.22).

To begin establishing these results, it is instructive first to compare the difference which the choice of communality imposes on the size of eigenvalues. The averoid and bi-factor data in Table 8 are from Harman (Reference 2, Table 9.21). The calculations were actually performed on different computers, as well, but Harman establishes (Reference 2, Table 9.23) that only very minor discrepancies can be associated with computer-to-computer differences. Major differences are due primarily to choice of communality.

It is strikingly evident that both with SMC communalities for the correlation matrix, and also with the average of eleven eigenvalues based on sample matrices from this matrix, the significance has all been concentrated into a single general factor. The similarity of this factor to the general factor based on other communality estimates is given in Table 9.

Table 8

Relation Between Communality Estimate and Eigenvalues of 24-Variable Matrix, With Sampling Error

	Communality Estimate				
Order	Averoid	Bi-Factor	SMC	Sample SMC	S. Dev. (10 d.f.)
1	7.63	7.66	7.66	7.55	.57
2	1.65	1.65	.38	.56	.33
3	1.17	1.18	.38	.34	.06
4	.90	.96	.29	.29	.07
5	.40	.42	.24	.26	.06
6	.35	.40	.23	.21	.06
7	.27	.31	.20	.18	.04
8	.25	.30	.18	.15	.04
9	.21	.23	.14	.12	.04
10	.14	.16	-.04	.10	.04
11	.07	.19	.00	.08	.04
12	.01	.05	-.01	.05	.04
13	.00	.03	-.08	.03	.06
14	-.04	-.01	-.08	.01	.06
15	-.08	-.07	-.10	-.02	.06
16	-.09	-.07	-.12	-.05	.07
17	-.13	-.09	-.15	-.08	.07
18	-.16	-.14	-.15	-.09	.08
19	-.18	-.16	-.17	-.12	.08
20	-.20	-.19	-.25	-.16	.05
21	-.24	-.21	-.27	-.20	.05
22	-.26	-.23	-.39	-.23	.05
23	-.31	-.27	-.43	-.28	.06
24	-.34	-.31	-.49	-.34	.08

It is clear that there is general agreement between this main factor calculated in the various ways. The first value in the "Average SMC" column for instance, .580, is the arithmetic average of 11 values, each in turn calculated from a sample of 145 observations. Those eleven values range from .472 to .669 with a standard deviation, as indicated, of .072. (There is a slight downward bias in these averages as calculated, because they have been scaled in the square metric to the eigenroot, and any averaging of the numbers ought to be done in the same way instead of arithmetically as here.)

Table 9

Comparison of a General Factor in 24-variable Matrix as Identified by Alternative Communality Estimates, With Sampling Error

Test	Averoid	Population SMC	Average SMC	S. Dev. (10 d.f.)
1	.596	.595	.580	.072
2	.373	.376	.367	.077
3	.418	.425	.421	.072
4	.484	.487	.451	.048
5	.689	.690	.658	.063
6	.685	.686	.684	.053
7	.676	.673	.663	.044
8	.676	.678	.647	.051
9	.693	.693	.675	.065
10	.466	.463	.463	.077
11	.557	.560	.550	.049
12	.466	.468	.485	.068
13	.601	.600	.593	.060
14	.425	.424	.423	.058
15	.391	.390	.364	.058
16	.506	.509	.498	.050
17	.465	.465	.462	.078
18	.520	.519	.511	.099
19	.444	.451	.483	.098
20	.616	.619	.631	.063
21	.595	.598	.597	.030
22	.612	.614	.615	.041
23	.690	.693	.686	.057
24	.651	.653	.656	.040
V	7.628	7.665	7.550	.58

So it would appear that the averoid-based general factor might have been hit upon by chance due to sampling the data and calculating SMC-based communalities. The surprising thing however is that the averoid-based factor is only one of four (see Harman, Reference 2, Table 9.22) whereas all the SMC-based samplings succeed in concentrating all the factor information into single factor.

A somewhat similar sampling relationship will come out if we compare communality estimates. Briefly, the averoid estimate for the first test was

.505, whereas the sampling provided 11 SMC estimates, varying from .428 to .666 and averaging .579. However, here we begin to discern the discrepancies which produce further factors in one case but not in the other. Nine of these eleven SMC estimates exceed the averoid estimate. In later tests all 11 SMC-based communalities exceed the corresponding averoid communality.

The result of this is that the factors of verbal rigidity, spatial, and memory, discerned by averoid-based communalities, are all absorbed into the general factor of the SMC-based communality. The 11 eigenvalues of the second factors have an average value of .556 as given in Table 8. Only two of the eleven exceed unity, and these two do not have the same sign pattern as any of the factors of Harman (Reference 2, Table 9.22).

As any objective test of the insignificance of the second factor, a sign test was made of the eigenvectors from the eleven samplings. If any significant weight, plus or minus, was in this second factor, then there would be a tendency for plus or minus signs to occur opposite that test in each of the eleven iterations. With 11 iterations and half of the weights minus, a non-parametric 1 per cent test would consist of 0, 1, 10, or 11 like signs corresponding to one of the 24 tests. None were observed. If we weaken the test to comprise 0, 1, 2, 9, 10, or 11 like signs, we have a six per cent test, and would expect to find 1.6 of the 24 tests with these sign compositions. In fact we found two, test 6 with nine minus signs and test 18 with two minus signs, just about as expected. Further, these signs are the opposite of what we would expect if we were measuring the verbal rigidity factor, the number two factor of the averoid analysis.

The conclusion here was quite unexpected but seems inescapable--the use of SMC communalities contributes far more-than expected to the parsimony with which the relationships in the Unitary Trait data can be described. It would seem that if further factors are to be discerned in the data, a much larger sample size must be employed.

Let us summarize the practical results of the foregoing analysis, as it touches the issues of uniqueness and sampling, i.e., how sure the investigator can be of his results.

1. The coefficient of variation of the largest root is 10 per cent.
2. The variance of the eigenvectors associated with this root is $5/Vn\sqrt{N}$.

3. One hundred to two hundred observations on each of the n variates should be enough. If more than 200 can be collected, split the data in half at random and run each half separately as well.

4. Use SMC communalities and make all factors beyond the first prove their existence before you accept them.

SECTION VII

APPLICATION OF FACTOR ANALYSIS

7.1 INTRODUCTION

Three applications of factor analysis are presented in this section. Subsection 7.2 contains explanations of how factors are interpreted for psychophysiological data. These examples should give greater insight into the interpretation problem in general. Factor analysis is presented purely as a representation technique in subsection 7.3.

7.2 FACTOR ANALYSIS OF PERSONAL HISTORY AND ANTHROPOMETRY DATA

Included in this section are two factor analytic studies which were performed on data collected in a psychophysiology experiment.* The first analysis is of personal history data ascertained from the subjects by a questionaire which contained approximately 150 items. Many of the variables were derived from more than one response, and some of the original items were deleted since they were discrete data points. After careful quantifying and scrutinizing, 41 variables were retained. Eighty-eight subjects were used. In this and the following study, the subjects were University of Dayton students.

The second study is concerned with 106 anthropometric measurements taken on 131 subjects in the same experiment. Unlike the personal history variables, this data set was already quantified. Variables included a number of heights, breadths, circumferences, and diameters.

In both studies, the principle components method was applied using unities as an estimate of communality. The number of eigenvalues greater than one was used as a completeness criterion, i.e., determining of the number of factors to be rotated. Varimax was the method of rotation employed for both.

The personal history data produced 14 factors. It is important to stress at this point that one must be extremely familiar with the

---

*These factor analyses were performed under Contract AF33(615)-1119 monitored for the U. S. Air Force by Major Victor H. Thaler, 6570th Aerospace Medical Research Laboratories. Wright-Patterson Air Force Base.

variables and what they represent, as well as the make-up of the subject group, before a meaningful interpretation of the factors can be made. In this case, all the factors could be identified conceptually.

The best approach to interpreting the various factors is to examine them one by one and note those variables which have the highest loadings. For example, in Table 10 it can be seen that Major Subject (.90), Educational Goals (-.74), and Vocational Plans (.88) have the highest loadings in factor 1, indicating that the factor is associated with educational-vocational plans. Since Vocational Plans and Major Subject were ranked from "academic" to "applied" in nature, and Educational Goals was rated in the direction of higher educational motivation, it would seem logical to conclude that the more applied the subject's interest, the more likely that advanced degrees (law, Master, Doctorate) are not desired. In addition, variables such as Home Address-Distance (.38) and Full Scale IQ (-.31) should be considered. This is where familiarity with the data is necessary. It was concluded here that I.Q. probably had a tendency to relate to higher educational goals and more academic interests. However, Home Address-Distance, which is the distance between the subject's home and the university, was thought to be a less universal value. The relationship is likely peculiar to this university because of its academic standards. Thus, this factor would be considered in terms of educational and vocational plans, with some degree of ability being associated.

The second factor is quite straightforward in interpreting. It is obviously related to socio-economic level. The variables Income of father (.71), Socio-economic Rating (-.77), Education of Father (.81), and Education of Mother (.68) load most highly, indicating a strong relationship between education and income. Note that Socio-economic Rating is negative in direction because of the scoring technique employed, i.e., the higher the level, the lower the score. Furthermore, the educational achievement of the father appears to be most important. Thus, the variables indicate the factor is a measure of socio-economic level.

The remaining factors may be interpreted in a similar manner. The interaction of the variables in some factors is more subtle and complicated, however, requiring a greater degree of insight.

The second example of a factor analytic study is that of 106 relatively homogeneous anthropometric measurements. As seen in Table 11, 17 factors were produced according to the eigenvalue criterion. The great number of high correlations between the variables is responsible for the dramatic reduction of variables to factors by approximately one-sixth.

Using the same interpretation technique as in the previous example, it can easily be observed which variables load most highly on each factor. Factor 1 in Table 11, for example, has a large number of variables loading at 0.6 and above. Both present and maximum weight of the subject load with a number of body breadths such as shoulder, chest, waist, and buttock, as well as various depths, circumferences, and skinfolds. Also showing some importance are the somatotypes. The first, or endomorphic, somatotype depicts the amount of softness and roundness characteristic in the subject's body, while the "G" (gynandromorphic) somatotype is concerned with the degree of femininity in the body. In addition, the third, or ectomorphic, somatotype loads at -0.45, and should be considered since it depicts the lean or frail body.

Consequently, the various items point toward a factor which explains general body size, but not height. The breadths, depths, and circumferences which load are those of the trunk, and do not include the extremities. The somatotypes must be considered, as they help clarify and confirm the nature of the factor. Obviously, the endomorphic and gynandromorphic body would have greater measurements on the pertinent variables, while the opposite would be true for the ectomorphic body. Thus, this factor can be labeled general trunk dimensions.

The second factor has its major loadings on measurements of height and, naturally, stature. Again, the loadings are extremely high, usually in the range of 0.80 to 0.95. To a lesser extent, hand

and foot dimensions appear in this factor. This is understandable, however, since taller people usually have longer hands and feet. Of equal importance in this factor is the third, or ectomorphic, somatotype with a 0.76 loading. The long, lean body which it represents is in line with the general nature of this factor. Therefore, it would probably be labeled as stature.

Continuing, the third factor may be interpreted as being relevant to grip strength or arm muscle. The highest loadings appear on the three grip strength variables, while minor loadings appear on the biceps and forearm measures. Although the later are about one-half the size of the grip strength variables, the fact that all other loadings are negligible and that the biceps and forearm dimensions logically relate to the strength necessitates their inclusion in interpretation.

By examination, factor 4 is a testicle factor, and factor 5 is a penis factor. Factor 6 is concerned with dimensions of the head, while factor 7 pertains more to facial measurements. The various measurements of the hands, wrist, and feet comprise the eighth factor. Similarily, the remaining factors may be defined by careful examination of significant loading and consideration of their conceptual importance.

While the above factors are all fairly clear because of the nature of the variables, this is not always the case. Therefore, it must be reiterated that without a complete understanding of the nature of the variables and the subject population, no meaningful interpretation can be made.

7.3 FUNCTION REPRESENTATION

A function may be represented in many different ways. The function "sine" has, for example, a Taylor series representation, a continued fraction representation, an infinite product representation, and a Chebyshev series representation.

The choice of method for representing the function depends on the purpose for which the representation is to be used. If it is desired to study a certain property of a function, a representation

is chosen which is known to highlight that class of properties. A Fourier series representation may be chosen when the frequency content of a function is of interest, for example. When the purpose includes evaluation of the function, properties of the representation such as speed and region of convergence help dictate the choice.

The properties of various classes of representation techniques have been the point of much interest in the history of mathematics, and probably the most studied class of representation techniques has been that of orthogonal function expansions. This is so because the properties of orthogonal function expansions have been found most desirable and useful in practice. However, a set of orthogonal functions is usually obtained in practice by solving differential equations. Thus, in order to have a set of orthogonal functions which reflect the properties of a class of functions, a differential equation must be associated with that class of functions.

Out of the proliferation of different orthogonal sequences such as the Legendre, Chebyshev, Laguerre, and Hermite polynomials came the unifying statement that all of these classical polynomials, $\phi_n(x)$, when multiplied by a particular weight function are solutions to the second order differential equation

$$G(x)y'' + \{2\,G'(x) - \phi_1(x)\}\,y' - \left\{\frac{n^2-n-2}{2}\,G''(x) + (n+1)\,\phi_1'(x)\right\}y = 0$$

where $y_n(x) = w(x)\phi_n(x)$. The effect of this statement was to provide a channel through which theory on one orthogonal sequence could be applied to another orthogonal sequence.

In the last five years physical scientists have shown interest in other methods which yield sets of orthogonal functions. The method about to be discussed may be characterized by an attempt to represent each member of a set of functions by a linear combination of nonlinear functions which span the space of possible given functions. The method obtains the basis functions by analysis of a symmetric, positive semidefinite matrix obtained from the given functions by various methods. Moreover, it is possible to

obtain a set of orthogonal basis functions which contribute maximally and in a decreasing manner to the total variance of the given functions.

In what follows the given functions $x_i(t)$ $(i = 1, 2, \cdots, n)$ will be represented discretely at N values of t by a vector with j^{th} component x_{ij}.

In factor analysis the given functions are first standardized by transforming to functions with zero mean and standard deviation of one. A correlation matrix with elements r_{ij} is then formed. Factor analysis provides very many methods for analyzing the correlation matrix including principal components.

The method of principal components depends on obtaining a representation of the transformed given function as

$$x_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{in}F_n$$

where it is assumed that the (F_k) are orthogonal functions. The method of principal components is based on the ability to spectrally resolve a linear symmetric operation into

$$R = \lambda_1 e_1 e_1^T + \lambda_2 e_2 e_2^T + \cdots + \lambda_n e_n e_n^T ,$$

where e_i is the normalized eigenvector corresponding to the eigenvalue λ_i of R. Then based on this spectral resolution of an operator, when

$$a_{ik} = \sqrt{\lambda_k}\, e_{ki}$$

is chosen,

$$a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{in}F_n$$

will indeed represent x_i since

$$xx^T = (AF)(AF)^T = AFF^TA^T = AA^T = R$$

Other methods of factor analysis make use of the full factor analysis model which includes unique factors or functions:

$$x_i = a_{i1}F_1 + a_{i2}F_2 + \cdots + a_{im}F_m + a_iU_i .$$

The purpose of this model which includes unique factors is to reduce to a minimum the number of factor functions $\{F_k\}$ which contribute to more than one given function x_i.

There are methods in factor analysis for obtaining a set of factor functions which are not orthogonal but oblique. These oblique factor functions are chosen so as to demonstrate the properties of the class of given functions in some way better than the orthogonal factor functions.

Orthogonal rather than oblique functions are usually used for function representation since they have such nice properties and are easy to handle. Indeed, going from orthogonal to oblique functions is like going from linear to nonlinear systems.

Nonlinear systems are, however, many times closer to reality. Just so, when it is desired to have the basis functions or factors represent concepts or actual causes of variance, oblique factors must be allowed since most conceptual causes of variance are related (therefore not independent or orthogonal.) Then let us consider the effect of these statements on the theory.

When the factor analysis of functions is stated as the problem of finding matrices A and F are such that

$$Z = AF \tag{1}$$

where A is the matrix ensemble of function vectors, A and F are underdetermined. There are an infinite number of matrices A and F which will satisfy Equation 1, just as there exist an infinite number of pairs of vectors which will span a two-dimensional space. In the principal components factor analysis discussed earlier, the condition of maximal, decreasing contributions to variance fixed the matrices and made the problem determinate.

If initially we have any A and F satisfying Equation 1, we may find others by "rotating" the given factors, i.e. by transforming each of the given factors by an orthogonal transformation matrix T. For example, in 2-space the factors may be rotated as shown in Figure 7.

In the analysis, the new oblique factors are derived by rotating a given (usually orthogonal) system to a new, preferred oblique system. It is postulated that a set of factors is more meaningful when each factor goes through a separate "cluster" of functions (when a group of functions are similar, their vector representations will be close to each other in space).

In attempting to find mathematical statements equivalent to this intuitive statement, most approaches reason as follows: When a factor passes through a cluster of functions, the coefficients of that factor for the nearby functions will be large while the coefficients of other factors for this cluster of functions will be small. This rationale has

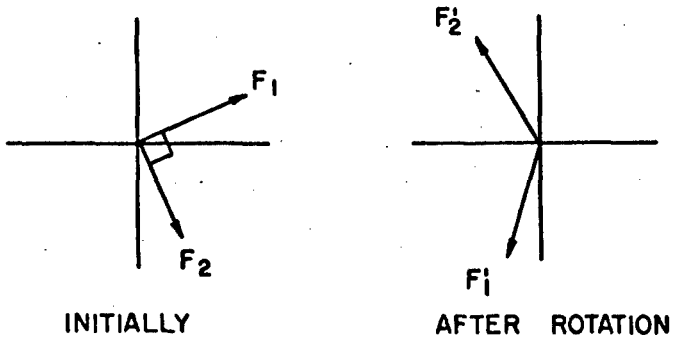


Figure 7. Rotated Factors

found its mathematical expression in the maximization or minimization of various functions of powers of various coefficients.

Representation using oblique factors is not quite so simple as in the orthogonal case since the coefficients are no longer Fourier coefficients. However the same method may be used to calculate the coefficients. For example, suppose we wish to expand a new function Z in terms of two known factors F_1 and F_2,

$$Z = a_1F_1 + a_2F_2 \tag{2}$$

If the factors are orthogonal, we find the Fourier coefficients by taking the inner product of both sides of Equation 2 with each of the factors. Thus,

$$\begin{aligned} (Z|F_1) &= a_1(F_1|F_1) + a_2(F_1|F_2) \\ (Z|F_2) &= a_1(F_1|F_2) + a_2(F_2|F_2) \end{aligned} \tag{3}$$

and when the factors are orthogonal,

$$(F_i|F_j) = \delta_{ij}$$

and

$$a_i = (Z|F_i).$$

However, with oblique factors

$$(F_i|F_j) \neq \delta_{ij} .$$

Thus the equations do not degenerate but must be solved simultaneously.

Table 10

Varimax Rotation of 41 Personal History Variables - 14 Factors

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Full scale IQ	-.31	.10	.14	.28	.09	.00	-.03	-.10	-.07	-.07	-.79*	-.07	.05	-.05
Age	.15	-.14	.03	-.07	-.09	.11	-.01	.82*	-.13	.06	.10	.06	.04	-.08
Home address-dist.	.38*	.11	.07	-.17	-.31	.05	-.08	.10	.25	.25	.26	-.29	.13	-.06
College rank	.21	.04	.04	.14	.13	.04	-.09	.84*	-.05	.09	.06	.00	-.11	-.06
Major subject	.90*	-.02	.07	-.04	-.04	-.03	.00	.07	-.02	-.16	.10	.13	.08	.07
Cur. avg. actual	-.14	-.10	-.06	.84*	.17	-.06	.20	.00	.07	-.00	.16	-.03	-.09	-.11
Cum. avg. actual	-.23	.11	-.02	.90*	.07	.01	.11	.03	-.05	-.05	.05	-.01	.01	-.03
X-curric. activity	.05	.04	-.05	.34	.11	-.11	-.03	.34	-.02	.13	.10	.45*	-.36*	.08
Age of father	.01	-.10	.04	-.07	-.89*	-.00	.09	.02	-.08	-.04	.16	.01	.03	.01
Income of father	.17	.71*	.20	.15	.13	.00	-.27	.10	-.06	.11	.10	.13	-.18	-.00
Subject employment	.02	.21	.13	.08	.03	-.26	-.11	.29	.01	-.10	-.01	-.19	-.65*	-.14
Describe father	-.03	-.01	-.03	-.08	-.04	.78*	-.14	.04	.13	-.16	.12	-.19	.11	.17
Age of mother	.17	.01	.01	-.03	-.88*	.01	.05	-.05	.06	-.07	-.11	-.01	-.01	.01
Describe mother	-.00	.11	.10	.12	.04	.83*	.09	.01	-.04	.11	-.04	.01	-.13	-.12
No. of siblings	.13	.10	-.12	.05	-.13	.03	-.11	-.11	.03	-.85*	-.01	.03	-.04	-.04
Sibling rank	.06	-.21	.01	-.05	-.59*	-.07	.02	-.08	-.04	-.67*	.03	.11	-.14	-.08
Illnesses	.02	-.05	.11	.06	.07	-.25	.06	.07	-.03	.00	-.05	-.27	.68*	-.15
Dating frequency	-.05	.12	-.01	-.07	-.07	-.12	-.01	.08	-.01	-.15	.09	.80*	-.02	.12
Visceral disorders	.11	-.10	.14	.01	-.03	.03	-.19	-.09	.76*	.02	-.03	.09	-.08	.07
Rank perf-IQ discp	-.06	-.09	.15	-.45*	.03	-.13	-.08	-.08	.07	.06	-.78*	-.07	-.04	.04
Acting out	.06	.01	.77*	.09	-.02	-.27	-.01	-.10	.27	-.04	.02	-.11	.07	.07
No. of operations	.24	-.09	-.30	.05	.17	-.01	.08	-.15	.17	.18	-.05	.52*	-.09	-.45*
Comb psycho symp	-.01	-.15	.38*	.03	.02	.01	.01	-.18	.71*	.10	-.00	-.07	.01	-.24
Educational goals	-.74*	-.05	.02	.21	.16	-.05	-.21	-.08	.04	-.07	-.11	.21	.02	.07
Age - concept	.07	-.15	-.25	.10	.11	-.03	.22	-.00	.58*	-.15	-.01	-.41*	.06	.05
Age 1st dated	-.04	.00	-.11	-.06	.17	-.08	.28	.31	.26	-.41*	-.15	-.25	.17	.18
Age 1st smoke	-.23	.05	-.14	-.15	-.03	.23	.08	.10	.50*	-.09	.17	.21	.33	.17
Age 1st drank	.12	-.03	.01	-.02	.05	.38*	.09	-.02	.11	-.15	.25	.06	-.04	.64*
Age 1st mstbt	-.13	.18	-.04	-.32	.10	-.07	.30	.47*	-.01	.01	.11	.08	-.17	.18
Size of metro area	-.02	.18	-.19	-.22	-.24	-.23	-.30	.03	.16	.48*	-.02	-.08	-.21	-.17
Age sex relat	-.02	.19	-.11	.17	-.16	-.21	.15	.54*	.02	-.09	-.34	.02	-.08	.41*
Age 1st drunk	.04	.13	-.14	-.12	.03	-.10	-.01	-.04	-.03	.13	-.13	.06	.04	.84*

Table 10 cont.

Varimax Rotation of 41 Personal History Variables - 14 Factors

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Car available(neg)	.16	.02	-.51*	.08	-.03	-.06	-.12	-.12	.12	-.05	.05	-.09	.54*	.15
No. auto accidents	.03	.08	.60*	-.19	.06	.20	-.14	.17	-.16	.03	-.13	.35*	-.12	-.08
No. of tickets	.13	-.02	.82*	-.08	-.08	.17	-.03	-.01	.11	.10	-.14	-.12	-.02	-.09
Vocational plans	.88*	.03	.06	-.20	-.04	-.03	.02	.15	.11	-.09	.07	.04	-.04	.07
Exp. income-10 yrs	-.18	.04	.05	-.13	.17	.01	-.85*	.01	.07	-.02	-.08	.03	-.06	-.02
High school rank	-.15	.01	-.05	.72*	-.11	.11	-.22	.02	.02	-.04	-.37*	-.02	.09	.02
Soc-econ rating	-.14	-.77*	-.07	-.11	-.06	.00	-.01	.00	.20	-.15	.10	-.14	.11	-.15
Education - father	-.08	.81*	-.09	-.04	.09	.10	.01	-.00	-.03	-.04	-.17	.03	.17	.01
Education - mother	-.11	.68*	-.09	-.16	-.11	-.02	.12	-.04	.00	-.15	.22	-.11	-.15	.02

Table 11

Rotated Factor Matrix

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Present weight	.90*	.34	.12	.03	-.[illegible]	.08	-.05	.09	-.06	.04	.03	.07	.03	-.06	-.06	.08	-.07
Maximum weight	.81*	.37*	.11	-.01	-.12	.08	-.07	.03	-.05	.05	.02	.09	.17	-.11	-.07	.19	-.09
Age at max weight	.20	-.08	-.02	-.07	-.07	.08	.03	.02	.04	.06	-.00	.01	-.84*	-.05	-.14	-.04	.04
Head circumference	.25	.07	-.08	.07	.08	.84*	-.14	-.01	-.07	-.03	-.04	.20	-.06	.03	-.00	.04	-.03
Head length	.16	.06	-.12	.01	.03	.84*	-.09	.02	-.05	-.10	-.07	-.22	-.06	.03	-.05	.15	.02
Head breadth	.08	.08	.16	.02	.04	.26	.03	-.03	-.05	.06	.03	.78*	.00	.00	.01	-.13	-.04
Bitra coronal arc	.24	.12	.17	.07	-.01	.63*	-.01	.01	.00	.06	.16	.44*	-.05	-.01	.13	-.10	.10
Bitra menton arc	.45*	.24	-.16	-.01	-.00	.33	-.08	.11	-.15	.19	.14	.21	.09	.09	-.01	-.06	-.32
Bitra min frnt arc	.30	.24	.03	.02	-.09	.63*	-.07	.14	.01	-.03	.07	.24	.03	-.03	.07	-.14	-.12
Total facial ht	.13	.12	.00	.06	.05	.15	-.93*	.05	.02	-.00	.02	.01	.02	-.04	.00	.06	.01
Lower facial ht	.09	.02	-.01	.05	.04	.14	-.85*	-.17	-.11	.09	-.01	-.13	.11	.09	-.11	-.02	-.04
Nasal length	-.06	.06	-.06	-.04	.04	-.09	-.68*	.42*	.14	-.09	-.01	.16	-.17	-.12	.08	.10	.04
Nasal breadth	.15	.10	.13	.12	.17	.39*	.18	.18	.00	.13	-.07	.09	.14	-.03	.25	.05	-.43*
Stature	.11	.93*	.07	.05	-.00	.04	-.08	.17	-.07	.06	.14	.12	.08	.03	.01	.07	-.01
Max stature	.11	.93*	.07	.06	-.00	.05	-.06	.18	-.08	.06	.13	.12	.09	.04	-.00	.07	-.02
Cervicale ht	.15	.94*	.04	.04	.01	-.00	-.06	.14	-.03	.06	.11	.09	.05	.02	.00	.06	.00
Acromial ht	.14	.94*	.01	.03	-.02	.01	-.02	.15	.02	.03	.20	.05	.02	-.01	.01	.03	-.02
Suprasternale ht	.17	.95*	.05	.04	-.02	-.01	-.02	.14	-.03	.05	.11	.10	.06	.01	-.01	.05	-.01
Nipple ht	.06	.96*	-.02	.04	-.01	-.03	-.04	.12	.00	.01	.16	.05	.03	-.01	-.02	.02	.02
10th rib ht	.12	.93*	.04	.04	-.01	-.03	-.05	.03	-.03	.02	.08	.03	-.00	-.05	-.00	-.04	-.02
waist ht I	-.11	.93*	.00	.08	.09	.06	-.00	.01	.07	.06	-.06	-.01	.06	-.07	.01	-.07	.03
Waist ht II	.07	.94*	.11	.09	.09	-.01	-.01	.04	-.04	.06	-.03	.07	.04	-.04	-.02	-.04	-.05
Iliac crest ht	.06	.96*	.05	.00	.11	.01	.03	-.00	-.00	.03	-.08	.00	.07	.01	-.00	.01	.00
Trochanteric ht	.01	.96*	.06	-.05	.10	.04	-.01	-.03	-.01	-.03	-.12	-.03	.04	.00	-.01	-.03	-.00
Symphysial ht	.03	.93*	.07	-.01	.12	.12	.02	-.02	.02	.06	-.07	-.04	-.02	-.15	-.02	-.07	.05
Penale ht	.04	.95*	.10	.02	.16	.06	.01	-.02	.02	.02	-.10	.01	.03	-.02	-.06	.03	.03
Gluteal furrow ht	.00	.95*	.07	.06	.09	.02	-.01	-.06	-.02	.02	-.08	-.04	.01	-.04	.03	-.06	.00
Kneecap super ht	.10	.90*	.09	-.03	.12	.12	.04	-.07	.01	.07	-.13	-.11	-.07	.01	.09	.02	-.05
Kneecap infer ht	.15	.86*	.09	-.04	.14	.08	.01	-.14	-.07	.08	-.18	-.15	-.09	.06	.07	.08	-.08
Tibiale ht	.12	.86*	.06	-.08	.10	.11	.05	-.07	-.04	.08	-.11	-.16	-.13	.01	.06	.07	-.09
Med malleolus ht	.18	.85	.05	.04	.01	-.07	.01	.10	.02	.84*	-.02	.06	-.01	.05	-.03	.13	-.05

Table 11 cont.

Rotated Factor Matrix

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Sphyrion ht	.14	.29	.05	-.06	.00	-.01	-.03	.09	-.03	.87*	.10	.03	-.06	-.05	-.11	-.05	.01
Radiale ht	.19	.88*	.05	.03	.02	-.01	-.03	.06	.05	.08	.33	.06	.01	.01	-.04	.04	-.00
Stylion ht	.21	.79*	.00	.09	.04	.01	-.02	.02	.03	.06	.48*	.04	-.03	-.01	-.15	.02	.02
Meta III ht	.17	.79*	.00	.10	-.00	.02	-.01	-.02	.05	.06	.47*	.06	.00	-.00	-.14	.05	.04
Sitting ht	.06	.61*	.04	.09	-.13	.06	-.07	.36*	-.20	.12	.45*	.20	.11	.11	.09	.18	-.01
Sitting acrom ht	.05	.52*	.15	.05	-.09	.01	.06	.37*	-.04	.13	.59*	.05	-.03	-.01	.10	.08	-.02
Biacromial dia	.42*	.34	.13	.13	.19	.03	-.02	.12	-.48*	.16	-.11	.26	.23	.03	-.06	.05	.03
Shoulder breadth	.77*	.17	.18	.03	.06	.09	-.05	.11	-.45*	.08	.03	.09	.01	-.08	-.08	-.02	-.05
Chest breadth I	.78*	.10	.12	-.12	.05	.06	-.07	.11	-.40*	-.06	-.07	.05	-.08	-.07	-.12	.12	.03
Chest breadth II	.76*	.21	.01	.02	.07	.17	-.01	.08	-.40*	-.01	-.10	.07	-.03	-.03	-.03	.07	.01
10th rib breadth	.90*	.11	.06	.03	-.03	.14	.03	.10	-.11	-.04	-.07	.02	-.09	-.01	.02	.03	.08
Waist breadth I	.92*	.09	-.05	.03	-.08	.13	.05	.09	.08	.11	.10	.08	-.02	-.07	.07	-.04	.14
Waist breadth II	.94*	.17	.02	.01	-.11	.13	.05	.01	.03	.04	-.03	.04	.02	.03	.07	.02	.09
Bicristale breadth	.74*	.34	.08	.12	.06	.12	.05	.12	.07	.10	-.00	.17	.00	.04	.13	.01	.05
Buttock breadth	.84*	.25	-.07	-.01	.00	-.11	.10	.06	.13	-.05	.15	.01	.04	.02	.02	.12	.11
Bitrochanteric br	.73*	.38*	-.08	.15	.03	.03	-.02	.21	.09	.20	-.08	.18	-.05	.03	.11	.05	-.11
1st grip strength	.08	.19	.86*	.07	-.04	-.02	.03	.05	-.08	-.01	-.07	.06	-.03	.06	-.00	.17	-.08
2nd grip strength	.12	.17	.87*	.01	-.02	-.02	.00	.10	-.04	.03	-.09	.04	.03	-.11	.03	-.06	.02
3rd grip strength	.15	.19	.85*	.04	.03	-.05	.01	.06	-.00	.06	.09	.10	.05	-.05	-.07	-.03	.08
Chest depth	.85*	.11	.16	-.03	.03	-.07	-.03	.05	-.10	.02	-.11	-.13	-.05	-.06	.06	-.07	.00
10th rib depth	.92*	.06	.04	-.01	-.07	-.08	-.04	-.05	-.10	.02	.05	-.07	-.04	-.07	.08	-.03	-.00
Waist depth I	.90*	-.04	-.03	-.03	-.13	.02	.01	.05	.03	.01	-.00	-.07	.02	-.14	.10	-.12	.16
Waist depth II	.92*	.08	-.00	-.01	-.15	-.03	-.03	-.01	-.00	-.01	-.04	-.06	.03	-.08	.10	-.06	.10
Buttock depth	.93*	.08	.03	-.07	-.08	.02	-.02	-.02	-.01	-.02	.05	-.05	-.03	-.09	-.02	-.07	-.04
Shoulder circ	.85*	.11	.12	.02	.04	.08	-.04	.11	-.36*	.08	-.03	.07	-.06	-.02	-.05	-.06	-.02
Chest circ	.88*	.13	.13	-.08	.03	-.00	-.08	.05	-.30	-.00	-.05	.01	-.06	-.06	-.07	.03	.03
Circ at 10th rib	.95*	.08	.05	.03	-.06	.04	-.02	.02	-.10	-.03	-.02	-.02	-.08	-.04	.03	-.02	.06
Waist circ I	.93*	.05	-.05	-.00	-.11	.12	.02	.08	.05	.03	.06	.02	-.03	-.12	.07	-.10	.15
Waist circ II	.97*	.12	-.01	-.01	-.13	.07	.01	-.01	.01	-.01	-.03	-.01	.01	-.00	.06	-.04	.09
Buttock circ	.93*	.22	-.00	.03	-.07	.04	-.08	.08	.03	.07	.02	.09	.03	-.04	-.04	.03	-.06
Upper thigh circ	.93*	.12	.08	.03	-.06	.09	-.02	-.02	.02	.04	.07	.06	-.00	-.05	-.09	.07	-.11
Mid thigh circ	.86*	.14	.13	.08	.03	.08	-.03	.00	-.00	-.02	.04	.10	.03	-.08	-.20	.10	-.16

Table 11 cont.

Rotated Factor Matrix

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Lower thigh circ 1	.89*	.08	.08	.04	.[illegible]	.01	-.07	.11	.05	.06	-.00	.08	.02	-.05	-.20	.14	-.12
Lower thigh circ 2	.85*	-.03	.07	.06	.07	.05	-.09	.13	.09	.08	.06	.11	.04	-.18	-.22	.13	-.10
Calf circ	.84*	.15	.13	.05	.02	.08	-.06	.09	-.04	.04	.01	.03	.06	-.03	-.20	.24	-.21
Ankle circ	.72*	.23	.19	.08	.12	.08	-.04	.29	-.01	.01	.02	.06	.11	.05	-.28	.17	-.13
Biceps circ I	.82*	-.02	.32	-.04	.07	.04	-.03	.09	-.23	.01	.13	-.00	-.12	.02	-.11	.04	-.15
Biceps circ II	.83*	-.03	.31	-.05	.04	.02	.00	.07	-.18	-.02	.15	.06	-.10	.03	-.14	.04	-.17
Forearm circ I	.73*	.06	.41*	.02	.13	.00	-.00	.23	-.15	.03	.15	-.07	-.05	-.03	-.19	.02	-.18
Forearm circ II	.73*	.09	.38*	-.03	.08	.02	-.01	.26	-.17	.05	.12	-.04	-.09	-.02	-.20	.02	-.18
Hand length	.24	.53*	.33	.00	.23	.10	-.12	.48*	-.01	.07	-.10	-.04	-.14	.03	.05	-.05	-.05
Hand breadth	.25	.43*	.28	-.05	.08	.08	-.04	.53*	.02	.17	.13	-.19	.01	-.04	-.09	-.19	-.07
Wrist circ	.64*	.26	.21	.02	.12	.01	.10	.42*	-.22	.17	.09	-.02	-.02	.09	-.14	-.01	-.15
Wrist breadth	.13	.29	.13	.00	.19	.03	.05	.68*	-.17	.13	.12	.01	-.03	-.01	-.12	.03	-.09
Foot length	.17	.54*	.10	.04	.21	.12	-.14	.55*	.03	-.02	-.03	-.00	.05	-.02	-.05	.05	.17
Foot breadth	.32	.43*	.11	.07	.14	.25	-.13	.42*	.16	.09	.07	.05	.29	-.10	-.28	-.08	-.01
Foot circ	.36*	.45*	.10	.13	.08	.30	-.05	.37*	.10	.12	.17	-.00	.21	-.12	-.32	-.03	-.05
Biceps skinfold	.74*	-.10	-.19	-.03	-.13	-.01	.02	.01	.11	-.01	-.01	.05	.12	.46*	-.02	-.04	-.02
Subscapular	.86*	-.08	-.11	.01	-.15	.04	.05	-.05	.02	.00	-.01	-.05	-.07	.30*	-.04	-.10	.08
Triceps skinfold	.73*	-.02	-.14	-.05	-.02	.04	.05	-.02	.23	.01	.01	.02	.01	.42*	-.03	-.04	-.18
Forearm skinfold	.69*	-.04	-.15	-.05	-.08	.03	.05	-.07	.08	-.02	.01	.04	18	.54*	.12	- 03	.04
Juxta nipple	.81*	-.18	-.15	.02	-.02	.09	.01	-.06	.09	-.01	.08	-.08	.03	.32	.04	-.17	.20
Mal x	.87*	-.04	-.09	-.07	-.17	-.04	.04	-.06	.07	-.03	-.00	-.03	-.02	.27	.05	-.04	.11
Juxta umbilicus	.85*	-.11	-.06	.01	-.10	.09	.01	-.13	.01	-.01	.04	-.15	-.01	.23	.07	-.16	.09
Mal u	.81*	-.11	-.07	.01	-.09	.11	-.02	-.15	.02	-.00	.08	-.14	-.13	.23	.09	-.15	.02
Thigh medial	.80*	-.06	-.08	-.09	.08	.14	.08	-.05	.09	.08	-.05	-.03	-.01	.30	.05	-.10	-.02
Calf medial	.76*	-.01	-.04	-.14	.09	.07	.08	-.04	.17	.01	.07	-.01	-.07	.35*	.03	.00	-.10
Acromion radiale	.03	.73*	-.06	-.04	-.09	.02	-.01	.31	-.04	-.06	-.28	-.04	-.03	-.09	.12	-.03	-.04
Radiale stylion	.12	.73*	.09	-.13	.11	.01	-.03	.22	-.10	-.01	-.27	.02	-.06	.03	.24	.04	-.14
1st somatotype	.81*	-.11	.04	.03	-.09	.10	-.09	-.00	.19	.06	.06	.05	.14	-.21	.04	.02	-.13
2nd somatotype	.24	-.18	.31	-.15	.13	.20	.02	.07	-.56*	-.07	.21	-.10	.04	-.24	-.22	.22	-.11
3rd somatotype	-.45*	.76*	-.14	.07	-.04	-.08	-.01	.14	.06	.04	.04	.09	-.09	.20	.08	-.05	.10
D somatotype	-.09	.19	-.06	.04	.08	.09	.04	-.11	.11	-.12	.01	.04	.16	-.04	.77*	-.06	-.05

Table 11 cont.

Rotated Factor Matrix

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
G somatotype	.68*	-.04	-.04	.15	-.01	.05	.03	-.22	.15	.01	.02	.14	.01	.08	.32	-.11	.03
Pl I	-.34	.22	.03	.03	.81*	.08	-.13	.05	-.05	-.03	.03	-.01	.12	-.05	-.04	.10	.02
Pl II	-.25	.18	.04	-.01	.84*	-.01	-.00	.04	.05	.05	-.03	-.02	.05	-.05	.05	.15	.05
P breadth at g	-.16	.17	-.05	.17	.56*	.03	-.06	.14	-.17	-.04	.01	-.28	-.03	.03	-.06	-.23	-.32
P breadth at r	.03	.13	-.16	.16	.77*	-.01	-.10	.07	-.07	-.01	-.03	.09	.02	.03	.07	-.09	-.23
P stretched	-.18	.26	.17	.04	.73*	.01	.15	.14	.01	.01	-.00	.09	-.08	-.00	.03	.15	.27
Right t length	.05	-.10	.01	.82*	.10	.01	.02	-.11	.06	.03	.11	.04	.03	.09	.12	.12	.02
Right t breadth	-.03	.10	.08	.77*	-.01	.05	-.00	.06	-.10	.12	-.01	-.01	-.08	-.18	-.03	-.01	.08
Right t depth	-.04	.03	.06	.40*	.05	-.03	-.36*	-.00	-.25	-.05	-.07	-.29	.20	-.07	-.06	.33	.18
Left t length	.08	.10	-.06	.79*	.04	.07	.01	.05	.09	-.19	.04	.02	-.02	.05	.03	.11	.01
Left t breadth	-.06	.14	.06	.77*	.10	.01	-.10	.02	-.01	.04	-.07	.01	.15	.00	-.08	-.04	-.21
Left t depth	-.05	.09	.07	.26	.16	.05	-.13	-.01	-.08	.06	.08	-.17	.03	-.03	-.07	.67*	-.00

Section VIII

RECOMMENDATIONS

Besides a survey of factor analysis, theory was extended in the areas of effects of the number of observations, sampling effects, interpretation of factors, and communality. There are other areas of factor analysis which are suggested to be further studied.

There are many multivariate analysis models which are closely related, such as, intrinsic analysis, Loève-Karhunen, latent structure analysis, and latent profile analysis models. A comparative study is needed to clarify similarities and differences of these models.

Factor analysis packages should be made more adaptive, i.e. more decisions could be made by the computer. For example, the number of factors for rotation, the grouping of variables, etc., as a matter of fact, the computer should handle the data up to the point of naming the factors. This would make factor analysis available to all scientists with a minimum effort on their part.

REFERENCES

1. L. Kelley. Comment on Wilson and Worcester's "Note on factor analysis". Psychometrika, 1951, 5, pp. 117-120.

2. H. Harman. Modern Factor Analysis. Chicago: University of Chicago Press, 1960.

3. M. G. Kendall. Rank Correlation Methods. Second Edition. London: Charles Griffin and Co., 1955.

4. A. L. Edwards. Statistical Analysis for Students in Psychology and Education. New York: Rinehart, 1946.

5. M. W. Richardson and F. M. Stalnaker. A note on the use of biserial r in test research. J. Gen. Psych., 1933, 8, pp. 463-465.

6. Ch. C. Peters and W. R. Van Voorhis. Statistical Procedures and Their Mathematical Bases. New York: McGraw-Hill, 1940.

7. K. Pearson. On a new method of determining correlation, when one variable is given by alternative and the other by multiple categories. Biometrika, 1910, 7.

8. B. H. Camp. The Mathematical Part of Elementary Statistics. Boston: Heath, 1931, p.309.

9. K. Pearson. On the correlation of characters not quantitatively measurable. Phil. Trans. A, 1900, CXCV, pp. 1-47 or in Trans. Roy. Soc. (London), Series A, 195, pp. 1-47.

10. H. W. Eber. Correlation program for up to 3000 variables. Program Write-Up, 1964, U. of Alabama, Birmingham, Ala.

11. L. Chesire, M. Saffir, and L. L. Thurstone. Computing Diagrams for the Tetrachoric Correlation Coefficient. Univ. Chicago Bookstore, 1933.

12. M. Hamilton. Nomogram for the tetrachoric correlation coefficient. Psychometrika, 1948, 13, pp. 259-271.

13. T. L. Kelley. Statistical Method. New York: MacMillan, 1923.

14. K. Pearson. On the theory of contingency and its relation to association and normal correlation. Draper's Company Research Memoirs 1, 1904 or in Mathematical Contributions to the Theory of Evolution. London: Cambridge University Press.

15. J. P. Guilford. Psychometric Methods. New York: McGraw-Hill, 1936.

16. G. U. Yule. On the methods of measuring the association between two variables. J. Roy. Stat. Soc., 1912, LXXV.

17. E. L. Thorndike. Mental and Social Measurements. Teachers College, Columbia University, 1913.

18. E. L. Thorndike. Your City. New York: Harcourt Brace, 1939.

19. E. Sones. A study of one hundred boys and girls, sixteen to eighteen years of age, who have left school and a similar group remaining in school. Master's Thesis, Pennsylvania State College, 1933.

20. M. R. Spiegel. Theory and Problems of Statistics. Shaum's Outline Series. New York: Schaum, 1961.

21. Health Sciences Computing Facility, University of California School of Medicine. BMD, Biomedical Computer Programs. Los Angeles: U.C.L.A. Student Bookstore, 1964.

22. F. R. Gantmacher. Applications of the Theory of Matrices. Interscience Publishers, 1959.

23. M. Ezekial and K. A. Fox. Methods of Correlation and Regression Analysis. New York: Wiley, 1959.

24. R. B. Cattell. Factor Analysis. New York: Harper, 1952.

25. L. L. Thurstone. Multiple Factor Analysis. Chicago: University of Chicago Press, 1947.

26. L. Guttman. 'Best possible' systematic estimates of communalities. Psychometrika , 1956, 21, pp. 273-286.

27. L. Guttman. Simple proofs of relations between the communality problem and multiple correlation. Psychometrika, 1957, 22, pp. 147-157.

28. W. Ledermann. On the rank of the reduced correlation matrix in multiple factor analysis. Psychometrika, 1937, 3, pp. 85-93.

29. P. H. DuBois et al. Partial Reduction Methods in Factor Analysis. AD 414860, 30 June 1963.

30. L. Guttman. Some necessary conditions for common factor analysis. Psychometrika, 1954, 19, pp. 149-161.

31. H. B. Heywood. On finite sequences of real numbers. Proc. Roy. Soc. Lon., 1931, 134, pp. 486-501.

32. H. S. Wilf. Mathematics for the Physical Sciences. New York: Wiley, 1962.

33. C. I. Mosier. Influence of chance error on simple structure: an empirical investigation of the effect of chance error and estimated communalities on simple structure in factorial analysis. Psychometrika, 1939, 4, pp. 33-44.

34. C. Burt. Tests of significance in factor analysis. British Journal of Psychology (Statistical Section), 1952, 5, pp. 109-133.

35. B. Fruchter. Introduction to Factor Analysis. New York: D. Van Nostrand Company, 1954.

36. R. R. Sokal. A comparison of five tests for completeness of factor extraction. Trans. of the Kansas Academy of Science, 1959, 62, No. 2.

37. L. R. Tucker. A matrix multiplier. Psychometrika, 1940, 5, pp. 289-294.

38. R. Blakey. A re-analysis of a test of the theory of two factors. Psychometrika, 1940, 5, pp. 121-136.

39. Q. McNemar. On the number of factors. Psychometrika, 1942, 7, pp. 9-18.

40. H. F. Kaiser. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 1960, 20, No. 1, pp. 141-151.

41. D. R. Saunders. Factor analysis III. Desiderata in experimental design.

42. K. Holzinger and H. Harman. Factor Analysis. Chicago: The University of Chicago Press, 1948.

43. T. L. Kelley. Essential traits of mental life. Harvard Studies in Education, 26. Cambridge, Mass: Harvard University Press, 1935.

44. L. L. Thurstone. The Vectors of Mind. Chicago: The University of Chicago Press, 1935.

45. R. A. Fisher and F. Yates. Statistical Tables for Biological, Agricultural and Medical Research. New York: Hafner Publishing Company, 1953.

46. C. Burt. The Distribution and Relations of Educational Abilities. London: P. S. King and Son, 1917, pp. 54-57.

47. D. N. Lawley. The estimation of factor loadings by the method of maximum likelihood. Proc. Roy. Soc. Edin. A , 1940, 60, pp. 64-82.

48. D. D. Rippe. Application of a large sampling criterion to some sampling problems in factor analysis. Psychometrika, 1953, 18, pp. 191-205.

49. Sten. Henrysson. The significance of factor loadings--Lawley's test examined by artificial samples. B. J. P. Stat. Sec., 1950, 3, pp. 159-165.

50. H. A. Reyburn and J. G. Taylor. Some factors of personality. Brit. J. of Psych., 1939, 30, pp. 151-165.

51. C. H. Coombs. A criterion for the significance of common factor variance. Psychometrika, 1941, 6, pp. 267-272.

52. F. Swineford. Some comparisons of the multiple-factor and the bi-factor methods of analysis. Psychometrika, 1941, 6, pp. 375-382.

53. P. G. Hoel. A significance test for minimum rank in factor analysis. Psychometrika, 1939, 4, pp. 245-253.

54. E. B. Wilson and J. Worcester. Note on factor analysis. Psychometrika, 1939, 4, pp. 133-148.

55. G. Young. Factor analysis and the index of clustering. Psychometrika, 1939, 4, pp. 201-208.

56. H. Hotelling. Analysis of a complex of statistical variables into principal components. J. E. P., 1933, 24, pp. 417-441, 498-520.

57. M. S. Bartlett. Tests of significance in factor analysis. B. J. P. Stat. Sec., 1950, 3, pp. 77-85.

58. P. E. Vernon et al. How many factors? Unpublished memorandum, 1949.

59. R. B. Cattell and J. L. Muerle. The MAXPLANE Program for factor rotation to oblique simple structure. Ed. Psychol. Measmt., 1960.

60. R. B. Cattell and M. J. Foster. The ROTOPLOT Program for multiple, single-plane, visually-guided rotation. Pre-publication copy from R. B. Cattell, 1962.

61. A. Hold. Statistical Theory with Engineering Applications. New York: John Wiley and Sons, 1952.

62. K. A. Brownlee. Statistical Theory and Methodology in Science and Engineering. New York: John Wiley and Sons, 1960.

63. H. M. Walker and J. Lev. Statistical Inference. New York: Henry Holt and Co., 1953.

64. A. Ralston and H. S. Wilf, Editors. Mathematical Methods for Digital Computers. New York: John Wiley and Sons, 1964.

65. B. C. HOW. Share General Program Library. Share Distribution Number 1321 PA.

66. H. W. Eber. Centroid factor analysis for 121-3,000 variables. Program Write-up, 1964, U. of Alabama, Birmingham, Ala.

Appendix I

COMPUTER PROGRAM WRITE-UPS

The factor analysis package presented in this appendix consists of four programs whose write-ups are contained in this appendix. The four program abstracts follow below in front of the write-ups themselves.

A. Factor Analysis Program

This program is a specialized version of the A70A program available from System Development Corp. which originated at the Harvard Statistical Laboratory. A factor matrix is computed using the Jacobi method. Input is restricted to a Pearsonian correlation matrix read from Fortran binary tape.

B. Factor Rotation Program

This program is computationally identical to the A26D program available from System Development Corporation. An orthogonal rotation is performed using the Kaiser Varimax criterion. Input is restricted to a Fortran binary tape prepared by the factor analysis program, SRL-FA1. An accuracy check is provided by computing and printing the differences between the original and the final communalities.

C. Oblimax Rotation Program

SRL-OB1 is a general purpose program which transforms the factor analysis model for a set of orthogonal factors to the model for a set of oblique ones, i.e., it rotates factors to a more meaningful oblique set.

Given an orthogonal factor pattern A on binary tape, the program uses the OBLIMAX criterion to find a transformation matrix Λ and reference structure matrix V such that

$$V = A\Lambda$$

as in Harman (Reference 2, p. 310). The heart of this rotation is the specialized version of an OBLIMAX rotation routine obtained from the University of Illinois.

Using V and Λ, other output forms of the oblique factor analysis model are then computed, in particular:

P - the new factor pattern

S - the factor structure

ϕ - the matrix of factor correlations

P and ϕ may be written on tape for further use.

D. Factor Scores Estimation Program

This program computes estimated factor scores using the equation

$$f = \phi A' R^{-1} Z$$

which is (16.2) in Harman (Reference 2, p.341). Input to the program consists of the correlation matrix R, the factor coefficient matrix A, the factor correlation matrix ϕ (if needed), and the raw scores. Output consists of the estimated factor scores (both listing and punched cards), as well as R^{-1}, test coefficients for standard scores and for raw scores, if desired.

A. Factor Analysis Program

CLASS: Self-Contained General Purpose Program

LANGUAGE: Fortran II

PURPOSE: To compute a factor matrix using the Jacobi method and write factor loadings on to binary tape.

RESTRICTIONS:

No. of variables	Maximum of 150
Input	Binary tape containing Pearsonian correlation matrix
Output	Available outputs in BCD mode: a) Correlation matrix b) Latent roots and vectors c) Factor loadings Output in binary mode: a) Factor loadings for input to factor rotation program

DESCRIPTION, USE & COMMENTS:

Tape Assignment

Logical Tape:

2	System BCD input tape.
3	System BCD output tape.
5	Correlation matrices in binary mode.
6	Factor loadings in binary mode.
9	Used for temporary storage of eigenvalues and eigenvectors.

Binary Input Tape Format

Record 1 (2 words)

Problem number and order of square matrix (N) - both in integer form.

Record 2 through N+1 (N words each)

One record for each row of correlation matrix (ones in diagonal).

Card Deck Preparation

Each problem to be run requires two data cards as follows:

A. Title Card

Col 1	PUNCH 1
Col 2-72	Any BCD information desired as page headings for printed output.

B. Problem Card

Col 1-5		Problem number used to locate proper matrix on input tape and to identify BCD output.
Col 6-8		Number of variables in this analysis.
Col 9-10		If all eigenvectors (and consequently factor loadings) are to be computed, leave these columns blank. Otherwise punch the reduced number of eigenvectors desired.
Col 11		
	= 1	The correlation matrix with communality adjustments is to be printed.
	= 0	This matrix is not to be printed.
Col 12		
	= 1	Latent roots and vectors are to be printed.
	= 0	Latent roots and vectors are not to be printed.
Col 13		Estimation of communalities
	= 1	Maximum row element.
	= 2	R^2 (square of multiple correlation coefficient of given variable with all other variables).
	= 3	Unities are retained.
	= 4	Image-covariance factor analysis. (Essentially R^2 with appropriate adjustment of the off diagonal elements to maintain the positive semi-definiteness of the matrix).

Col 14

= 1 Factor loadings are to be written on logical tape 6 for input to factor rotation program.

= 0 Factor loadings are not to be written.

C. Finish Card

Col 1-6 Punch FINISH

D. Blank Card

Note: Cards C and D must follow the Problem Card for the final problem.

ROUTINES USED IN SRL-FA1

REWV This routine is provided to rewind and unload tapes.

FMLEV This routine determines the eigenvalues and eigenvectors of a symmetric matrix. It is one of several eigenvector routines which are available from SHARE.

B. Factor Rotation Program

CLASS: Self-contained General-Purpose Program

LANGUAGE: Fortran II

PURPOSE: To perform an orthogonal rotation using the Kaiser Varimax criterion.

RESTRICTIONS:

Number of variables	250 maximum
Number of factors	50 maximum

DESCRIPTION, USE & COMMENTS:

Tape Assignment

Logical Tape:

2	System BCD input tape.
3	System BCD output tape.
6	Binary input of factor loadings from SRL-FA1

Card Deck Preparation

A. Problem Card

Each problem to be run requires a single control card as follows:

Col 1-5	Problem number used to locate proper factor loadings on input tape and to identify BCD output.
Col 6-8	Number of variables.
Col 9-10	Number of factors to be read from tape and rotated.

B. Finish Card

A series of problem cards is followed by the following card to signify that all problems desired have been run:

Col 1-5	Punch 09999

SUBROUTINES:

REWV	This routine is employed to rewind and unload the binary input tape.

C. Oblimax Rotation Program

CLASS: Self-Contained General Purpose Program

LANGUAGE: Fortran II

PURPOSE: To rotate orthogonal factors to a set of oblique factors using the OBLIMAX criterion and to compute various output forms of the model.

RESTRICTIONS:

Matrix size	No. of variables plus no. of factors $\leq$ 130.
Input	Binary tape containing orthogonal factor pattern.
Output	Available in BCD mode 1) Transformation matrix 2) Reference structure 3) Reference vector correlations 4) Reciprocals and inverses of elements of diagonal matrix 5) (Primary) factor pattern 6) (Primary) factor correlations 7) (Primary) factor structure Binary mode 1) (Primary) factor correlations for input to second order factor analysis 2) (Primary) factor pattern

DESCRIPTION, USE & COMMENTS:

Tape Assignment

Logical Tape:	
2	System BCD input tape.
3	System BCD output tape.
6	Factor pattern input and output in binary mode.
5	Factor correlation output in binary mode.

Binary Tape Input Format

Record 1 (3 words)

Problem number, no. variables (NVAR), no. factors (NFAC).

Record 2 through NFAC + 1 (1 + NVAR words each)

One record for each column of factor pattern.

Each record contains one dummy word followed by NVAR loadings.

Record NFAC + 2 (3 A6 words for each variable) variable names

Binary Tape Output Format

Factor Correlations Tape 5

Record 1 (4 words)

Problem number, 2 dummy variables, no. factors (NFAC).

Record 2 through NFAC + 1

One record for each row of the correlation matrix.

Record NVAR + 2 factor names

The values 1 through NFAC are set up in the 3 A6 words for factors 1 through NFAC respectively.

Factor Pattern Tape A6

Same as binary input format.

Card Deck Preparation

Each problem to be run requires two data cards as follows:

A. Problem Card

Col 1-5	Problem number used to locate proper matrix on input tape and to identify printed output.
Col 6-8	No. variables in this pattern.
Col 9-10	No. factors to be rotated.
Col 12-14	BCD output parameter. Form the sum for desired output: 100 - V, reference structure 010 - Ø, factor correlations 001 - S, factor structure blank - P, factor pattern (always given).

Other options
200 - all output listed on page [2]. Modified format where available.
400 - all output listed on page [2]. Modified format (PWRITE) and original output (six-place accuracy).

Col 16-20

JOB NUMBER to be written with factor correlation on tape 5 matrix for use in locating it.

blank - correlation matrix will not be written on tape 5.

Col 22-25

JOB NUMBER to be written with factor pattern on tape 6.

blank - factor pattern will not be written on tape 6.

Col 28

Leave blank unless starting new binary tape
1 - start new tape on logical unit 5
2 - start new tape on logical unit 6
3 - start both new.

Col 30

1 - the variables are to be normalized during rotation (made of length 1 in common factor space to insure that structure values indicate angular closeness of fit).

blank - communality of variables left unchanged.

Col 31-34

F4.2 conversion
* appears on BCD (printed) output beside values whose absolute value is greater than or equal to this number.

blank - the value 0.35 is used.

2.0 - no *'s; three-place accuracy.

B. Title Card

Col 1-78

Any BCD information will be written at the top of every page.

C. Finish Card — Last card in deck.

Col 1-5 — 99999

ROUTINES USED IN SRL-OBl

PREAD — Reads the binary input factor pattern.

PWRITE — Prints the selected BCD output.

RWRITE — Writes $\emptyset$ and P on tape.

OBMAX — Performs the rotation and calculates the output.

GENINV — A routine for symmetric matrix inversion.

MATHEMATICAL NOTES

1. A mathematical explanation of the OBLIMAX rotation process may be found in Harman (Reference 2, p. 310-319). The treatment is sketchy in one respect and the following extension may be helpful for a complete understanding of this program. Harman's terminology will be used.

OBLIMAX tries to maximize a function on the elements of the Reference Vector Structure matrix by an iterative process which successively maximizes the function in each of the planes formed by each pair of reference vectors. Although the end result is a transformation from orthogonal vectors to oblique ones, before the end of the first pass we must consider planes formed by, and transformations on, oblique vectors.

Therefore let us examine the general case of factor rotation in the plane of the j^{th} and k^{th} vectors. We are looking for a transformation which maximizes a function on the values v'_{ij}, $i = 1, \ldots, n$, (see 15.4, Harman) where

$$v'_{ij} = \lambda_{11} v_{ij} + \lambda_{21} v_{ik} \qquad (1)$$

and where v_{ij} is the correlation between variable i and reference vector j, or $(Z_i \Lambda_j)$. But note that (1) does not define unique λ_{11} and λ_{21} but a "line" of them. This is a reflection of the fact that in the oblique case structure alone does not determine the factor analysis description of common factor space. Such a description requires two of the several related matrices. Thus in the plane we must use λ_{11} and λ_{21} to transform one more set of values, i.e., find another equation consistant with (1). The most practical solution is to let λ_{11} and λ_{21} transform the old reference vectors into the new one:

$$\Lambda'_j = \lambda_{11} \Lambda_j + \lambda_{21} \Lambda_k$$

As Λ_j' is not yet known we may only deduce by squaring

$$(\Lambda_j'|\Lambda_j') = 1 = \lambda_{11}^2 + \lambda_{21}^2 + 2\lambda_{11}\lambda_{21}(\Lambda_k|\Lambda_j) . \qquad (2)$$

Likewise as v_{ij}' is not yet known we must find the set of values $(\lambda_{11},\lambda_{21})$ for which

$$k = \frac{\sum v_{ij}'^4}{\sum v_{ij}'^2 2} = \frac{\sum (\lambda_{11}v_{ij} + \lambda_{21}v_{ik})^4}{\sum (\lambda_{11}v_{ij} \quad \lambda_{21}v_{ik})^2 2}$$

is a maximum. For convenience let

$$x = \frac{\lambda_{21}}{\lambda_{11}} \qquad (3)$$

Then $\lambda_{11}v_{ij} + \lambda_{21}v_{ik} = (v_{ij} + v_{jk}x)\lambda_{21}$ and we may simply solve

$$\max\ k = \frac{\sum (v_{ij} + v_{ik}x)^4\lambda_{21}^4}{\lambda_{21}^2 \sum (v_{ij} + v_{ik}x)^2 2}$$

for x because the λ_{21}'s factor out and cancell.
Now combining (2) and (3)

$$1 = \lambda_{11}^2 + x^2\lambda_{11}^2 + 2\lambda_{11}x(\Lambda_j|\Lambda_k)$$

and solving for λ_{11} and λ_{21} yields

$$\lambda_{11} = \frac{1}{\sqrt{x^2 + 2x(\Lambda_j|\Lambda_k) + 1}}$$

$$\lambda_{21} = \frac{x}{\sqrt{x^2 + 2x(\Lambda_j|\Lambda_k) + 1}} \qquad (4)$$

OBLIMAX, after finding x, finds the denominator in (4) by normalizing the vector $\Lambda_j + x\Lambda_k$ where Λ_k and Λ_j are the k^{th} and j^{th} columns of the "total" transformation matrix Λ (from the original set of orthogonal reference vectors to the set of oblique ones). An "updated" transformation matrix may then be generated by

$$\Lambda_j' = \lambda_{11}\Lambda_j + \lambda_{21}\Lambda_k .$$

Of course OBLIMAX provides two values for x and the same procedures are applied, using the other value to find the new k^{th} column of Λ as well, thus rotating both the k^{th} and j^{th} reference axes.

2. When the rotation is finished, OBLIMAX has produced a transformation Λ with a double use: it transforms the initial orthogonal reference structure (equivalent to the factor pattern in the orthogonal case) to an oblique reference structure

$$V = A\Lambda$$

and it transforms the initial factors into the new oblique reference vectors. Hence Λ contains in its columns the direction cosigns of the new reference axes, using the initial set as an orthogonal basis.

The program then computes the matrix of correlations ψ between reference vectors

$$\psi = \Lambda'\Lambda .$$

The transformation Λ from the orthogonal factors to the reference vectors and a hypothetical transformation T from the original factors to the new set of oblique factors are related by

$$D = T'\Lambda \qquad (5)$$

where D is the diagonal matrix of the scaler products, or correlations, between vectors T_p and Λ_p ($p = 1, \ldots, n$). Because Λ_p is defined as the vector normal to the hyperplane of all factors T_q, $q \neq p$, it is uncorrelated with every factor except T_p, and hence D is diagonal.

From (5)

$$T' = D\Lambda^{-1}$$

tells us that T' may be calculated from Λ^{-1} by normalizing its rows, since the rows of T' are normalized and left multiplication by a diagonal matrix is equivalent to multiplying each j^{th} row by

the element d_j. To normalize Λ^{-1} by rows we may multiply each row by the reciprical of the square root of the diagonal elements of $\Lambda^{-1}(\Lambda^{-1})'$.

$$\Lambda^{-1}(\Lambda^{-1})' = \Lambda^{-1}(\Lambda')^{-1} = (\Lambda'\Lambda)^{-1} = \psi^{-1} \qquad (6)$$

So OBLIMAX simply inverts ψ and finds the elements of D as explained above. It then finds the oblique factor pattern P and matrix of correlations between factors Φ by the formulas derived here.

$$\Phi = T'T = D\Lambda^{-1}(D\Lambda^{-1})' = D\Lambda^{-1}(\Lambda^{-1})'D' = D(\Lambda'\Lambda)^{-1}D$$

$$\Phi = D\psi^{-1}D$$

$$P = S\Phi^{-1} = (AT)\Phi^{-1} = AT(T'T)^{-1} = ATT^{-1}(T')^{-1} = A(T')^{-1}$$

$$V = A\Lambda$$

$$A = V\Lambda^{-1}$$

$$P = V\Lambda^{-1}(T')^{-1} = V(T'\Lambda)^{-1} = VD^{-1}$$

$$P = VD^{-1}$$

Finally, the factor structure may be computed

$$S = P\Phi .$$

3. NOTE: the program has an option to normalize variables in common factor spaces during rotation. Then the OBLIMAX function is maximized on

$$v_{ij} = \frac{(Z_i|\Lambda_j')}{|Z_i|}, \quad i = 1, \ldots, n$$

instead of

$$v_{ij} = (Z_i|\Lambda_j'), \quad i = 1, \ldots, n .$$

This change eliminates the effects of differing variable communalities, making angular closeness of fit the determining factor.

D. Factor Scores Estimation Program

CLASS: Self-Contained General Purpose Program

LANGUAGE: Fortran II

PURPOSE: To estimate factor scores using the equation

$$f = \phi A' R^{-1} Z$$

where A is the n x m matrix of common factor coefficients, R is the n x n matrix of correlations (unity in the diagonal), Z is the n x N matrix of standardized scores, and f is the m x N matrix of estimated factor scores. ϕ is an m x m matrix of factor correlations; it is not used in orthogonal solutions.

RESTRICTIONS:

No. of variables	Maximum 90
No. of subjects	Maximum 90

DESCRIPTION, USE & COMMENTS:

Tape Assignment

Logical Tape:	
2	System BCD input tape.
3	System BCD output tape.
6	Original correlations and factor correlations in binary mode.
7	Factor loadings in binary mode.
11	Raw data in binary mode.
15	Factor scores in BCD mode for punching.

Card Deck Preparation

Each run requires the following cards:

A. Title Card

Col 1-78	Any BCD information desired as page headings.

B. Problem Card 1

Col 1- 5	Problem number used to locate original correlations on input tape and identify BCD output.
Col 6-10	Number used to find factor loadings.
Col 11-15	If an oblique solution, enter the number which identifies the factor correlations on tape. If orthogonal leave blank.
Col 16-18	Number of factors for which factor scores are to be computed. Must be equal to or less than the number on tape.
Col 19- 21	Number of tape batteries of raw data making up variable set.
Col 22	
=1	R inverse is printed.
=0	R inverse is not printed.
Col 23	
=1	Test coefficients (standard scores) are printed.
=0	Not printed.
Col 24	
=1	Test coefficients (raw scores) are printed.
=0	Not printed.

C. Problem Card 2

Col 1-78	File identification in 6 col fields as indicated in Col 19-21 of previous card. Maximum of 13.

Repeat above cards for each job.

D. Finish Card

Col 1-6	Punch FINISH

E. Blank Card
F. Blank Card

ROUTINES USED IN SRL-FS:

INVERT	Computes R inverse
TION 11	Positions tape 11 at correct raw data file.
LSHFT	Shifts integer numbers into FORTRAN II format since tape 11 is written in FORTRAN IV.

Appendix II

TIME FUNCTIONS OF COMPUTATION

In this appendix we present a compilation of data which will facilitate estimation of computation times on various computers. The factor analysis techniques may be described in terms of the basic matrix operations, sum, product, inversion, and eigenvalue and eigenvector computation. The following table gives the computation time where μ, δ, and α are the multiplication, division, and addition times, respectively, for a given computer.

1. Computation of all eigenvalues and eigenvectors of matrix $A_{N\times N}$ by the Jacobi method (Reference 64):

$$T = 10\,N^3\,\mu + 20\,N^3\,\alpha$$

2. Inversion of a symmetric matrix $A_{N\times N}$ by bordering:

$$T = N^2(N-1)\mu + \frac{1}{3}N(N^2+2)\delta + \frac{1}{6}(N-1)(4N^2 - N + 15)\alpha$$

3. Multiplication of $A_{N\times M} \cdot B_{M\times P}$:

$$T = NPM\mu + NP(M-1)\alpha$$

4. Addition of $A_{N\times M} + B_{N\times M}$:

$$T = NM\alpha$$

5. Computation of all eigenvalues and eigenvectors of matrix $A_{N\times N}$ by the Householder-Ortega-Wilkinson method:

$$T = .00162N^2,$$

where T is the time in minutes on the IBM 704 computer. Application of this equation to another computer will require multiplication by a scale factor which reflects the ratio of

speed of the other computer compared to the IBM 704. This empirical equation was derived by least squares methods from data given in Reference 65 .

Appendix III

DESIGN OF A FACTOR ANALYSIS

1. Data Collection

There are features and properties of factor analysis which are learned from experience by users, but which are rarely written into textbooks on the subject. The purpose of this appendix will be to touch on some of these features.

Factor analysis is performed on data which, geometrically speaking, consists of N points each situated in n-dimensional space. The purpose of factor analysis is to describe the shape of the set of N point as comprehensively and briefly as possible through mathematical shorthand.

In this framework, some of the shortcomings of factor analysis can be described. In the first place, only the correlation between pairs of variables is used to describe the raw data. This constitutes a drastic reduction of the data into very few numbers. If N=200 in ordinary three dimensional space, then forming the correlations involves reducing 600 numbers into only 3 numbers. Factor analysis reverses this process, and from these 3 numbers manufactures 3 characteristic roots and 9 characteristic vector elements. Evidently the entire process depends on how adequately all the information in the 600 numbers can be condensed and contained in only 3 numbers.

These three numbers are the product-moment correlations between the variables. These are, to begin with, pairwise expressions. They take each pair of variables, 1 and 2, 2 and 3, 1 and 3, and presume to describe in one numerical quantity what the relationship is between each pair. It becomes clear that much of the important information about the shape of the set of N points may be lost. It will depend, of course, on the shape of the set. The implication is clear. Look at the data out of which the correlations are being calculated. It is not feasible to try to make 3 dimensional sketches, and besides, there will usually be far more than just three variables. From n variables there will be n(n-1)/2 different pairs of variables, and the same number of correlation coefficients. Even plotting out all these graphs will be a major job, and for practical purposes it will

be sufficient to plot only a portion of all the N points defined by the available data.

What should the researcher be looking for? There are three basic danger signs to look for.

a. Outliers: data points which don't belong in the set, either because of incorrect collection or copying of data or irrelevant data.

b. Multiple populations: data points will be found to form two clusters in some graphs, in which case a different factor analysis for each cluster will be necessary. In practical problems the difference will be due to some observable fact such as differences of sex, production line, experimental technique, etc., which was initially ignored because it was considered unimportant for purposes of this analysis.

A more difficult danger sign in this connection is the presence of multiple populations not separated by distance. The only way to spot this is to go back to the raw data whenever a graph is found whose points follow an X, Y, or V shaped pattern. The purpose will be to see whether points on the one leg of the V have any other feature in common. No rigid rules can be given here. The picture will never be as clear-cut as is suggested here, and only experience can guide the researcher into those habits and practices of data examination which ferret out suspicious weaknesses in the original design of data collection.

c. Curvilinearity of data: the product-moment correlation coefficient measures the strength of relationship between two variables only if that relationship is linear. If the graph of the data plots into the shape of a C or S , then the whole projected factor analysis should be stalled at least temporarily until a statistician can be shown the data. The various options which might be recommended by him at this point go beyond the scope of this study.

These are the major danger signals. There are others, such as heteroscedosticity (data points pinched together at some places on the graph and spread out at others), but here again the investigator should be guided by the general warning--if anything looks suspicious, ask about it.

The actual calculation of the product-moment correlation coefficient is described in Section 2, and will not be spelled out here. What is not so frequently described, and often badly needed, is advice about avoiding biases due to improper data collection.

The framework for such a description must begin with the classical distinction between population and sample. Ideally, we might want to construct, for each pair of variables, the population correlation coefficient. For practical purposes this would be unwise in most cases. If only because the labor, editing, and error control would be so demanding, we would be led to sample.

It is in defining this sample that bias is apt to enter, particularly since any investigator is initially prone to the temptation to feel that a big correlation is a good correlation. It is only with experience that an investigator comes to accept the statistical standard that the population correlation, or an unbiased approximation to it, is the only good correlation. To bias a correlation coefficient, it is necessary only to remove a few observations from the middle of the set of observations, and since most observations will be in the middle in any case, such a removal will not seem particularly unprofessional.

The professional standard which will be adhered to is the criterion of random sampling--each data set should have the same chance of having its data incorporated into the computations as any other data set. Whether this is accomplished by strict random sampling, systematic sampling, or cluster sampling is irrelevant here--it is the criterion which must be strictly adhered to if the sample correlation coefficient is to contain all the information that it can about the population coefficient.

Another important issue in connection with sampling is that of sample size. How large a sample ought one to take? Here again no attempt will be made to repeat the technical approach taken by most textbooks, but to deal in terms of insights. There is a

popular feeling that something called a "law of averages" exists. Among non-professional people, this law exists as a feeling that something ought to happen, and few people would dare to try to formulate the "law of averages" specifically, in the sense that they might formulate the law of gravity or Archimedes principle explicitly. Part of the reason is that certain key concepts such as variance are not part of common knowledge, and that an explicit formulation of the law of averages requires this concept.

The best that can be done to formulate the law of averages without using the idea of variance is to say that an average (height, weight, etc.) will be "improved" if it is based on more and more observations. When the law is formulated explicitly it appears that this "improvement" is subject to another law, commonly referred to as the "law of diminishing returns". More specifically, it says that bringing in more observations does improve the accuracy of an average, but that the hundredth sample does not contribute as much as the tenth observation, and the thousandth observation contributes even less.

These laws also apply to estimating a product-moment correlation coefficient. The larger the sample, the better will be the coefficient probably. However, successive samples contribute less and less to the goodness of the estimate. (These are crude statements only of the situation, and are intended to be only a first approximation to the kind of formulation which would satisfy a professional statistician.)

The actual rate of convergence of the sample correlation coefficient to its true population value cannot be simply described, since it depends on what the true value is. If the true correlation is high, only a small sample is needed, whereas if it is near zero a large sample will be required. Since in factor analysis the one sample we draw will have to serve for estimating many correlations, it seems desirable to concentrate only on those correlations where we are likely to be in trouble, that is, cases of zero correlation in the population.

2. Basic Requirements for a Factor Analysis

The first issue facing the investigator will be that of deciding

whether factor analysis is at all relevant to the problem facing him. Consultation with a professional factor analyst of course is the best advice that can be given, but in certain situations it may be safe to proceed with no more than the guidance given here.

Factor analysis was first employed in personality testing and intelligence testing, and the conditions required for using it can be described with reference to an analogous situation from psychology. The reader can then decide for himself whether these conditions apply to the experimental data he is faced with--whether from an assembly line, an electrocardiograph or a radar or radio signal full of unwanted noise.

First, all the variables must be results rather than causes. They must be analogous to school examination results from different subjects--mathematics, physics, music. If any of the variables are causes--such as parents' I.Q. or education, pre-school play habits, etc.--and the purpose of the study is to find the relation between causes and effects, then factor analysis is not the proper technique.

Secondly, the investigator should ask himself whether the kind of answer provided by factor analysis will be at all relevant to the question he is posing as he looks at the data. That answer, in the school analogy, will be something to this effect: there is one factor with high weighting on all subjects, a second with high weighting on mathematics and physics and negative weighting on music appreciation. It will be up to the investigator to discover or decide that the first factor is general intelligence and the second is scientific aptitude. But it must be kept in mind that this kind of answer may not be what is really wanted. If the investigator is really interested in deciding who should be admitted to college, or whether boys differ from girls in scientific ability, then he should look for new or different analytic techniques. Factor analysis should never be undertaken solely because the data are in the proper form for factor analyzing. Any data processing technique such as factor analysis should be treated as relevant or irrelevant depending on what problem is being posed, what hypothesis is being

tested, why the data are being collected in the first place.

Assuming that the two foregoing conditions, absence of causal data and relevance of factor analysis, have been met, we may now turn to issues of proper data collection. Analysis is bound to be better if good data are collected, and irrelevant data rejected. Good data will have the following characteristics:

a. Completeness: Each data set will contain one observation on each of the variables incorporated. This condition is not absolutely essential, but it eases the computation burden considerably, whether calculations are performed on desk calculators or electronic computers.

A trivial and an unrealistic example will show how one must proceed. Suppose the letter x represents a missing observation, and the data consists of six data sets each of five variates, namely (2, 1, x, 3, x), (x, 3, 2, x, 7), (3, 2, 5, x, 1), (2, x, 3, 1, x), (3, 4, x, 1, x) and (4, x, x, 3, 5). To form the correlation between the first two variates, we can use only the first, third and fifth data set, since only these contain data on both of these variates. However, note that we will encounter difficulties in calculating the correlation between the third and fourth variate, since only the fourth data set contains observations on both variables, and a correlation cannot be computed from one such pair. The investigator must watch for this kind of situation. One other condition must be met before we can proceed to accept in this way numerical material containing missing data. That is, there must be no relationship between the magnitude of the missing numbers and the fact of their being missing. If the missing numbers are all unusually large, or unusually small, then nothing at all can be done with the data.

b. Relevance: Factor analysis will be much improved if the investigator has some intelligent suspicions as to what factors might emerge. In such a situation, the most desirable thing is to choose variables which will yield the factor if it exists. Thus if a range of scientific ability is expected as a factor, then we should incorporate variables on physics, chemistry, art and music, with the hope that one factor will have positive weight on the first two and negative weights on the

last two, either before or after rotation. Of course, factor weightings are non-directional and the signs of the weights may be reversed, yielding in effect an anti-scientific factor. This will be due to the arbitrariness of the calculations, and the investigator can change all the signs before publishing the results, in order to be able to provide psychologically meaningful names for the factors. Even the major factor, the general intelligence one in any examination test data, may have negative weightings on all the items and thus measure general stupidity instead of general intelligence. Each factor is a dimension, such as stupidity-intelligence, and we may refer to the factor by either pole of the dimension, or by both if the opposite polarity is not clear from the context. Guilford has suggested collecting three variables for each factor suspected to exist, and this number three should be regarded as a minimum.

c. Factorial simplicity: Ideally, each variable should contribute to a very significant degree to only one underlying factor, otherwise the factorial structure of the data is rendered very complex, and even rotation will fail to clarify the factors into meaningful psychological entities. The foregoing is formulated in terms of the school grade analogy, but the situation is the same in any field of investigation.

d. Unbiasedness: The data must, insofar as possible, constitute a random sample of the population whose factor structure we are trying to describe. That is, each element in the population should have the same opportunity as any other element to be incorporated into the sample.

e. Linearity: Raw data are not used directly in a factor analysis. Rather, the relationship between all possible pairs of variates, as measured by the product-moment correlation coefficient

$$r_{ij} = \frac{\sum(X_i-\bar{X}_i)(X_j-\bar{X}_j)}{\sqrt{\sum(X_i-\bar{X}_i)^2\sum(X_j-\bar{X}_j)^2}}$$

is employed. Other measures of correlation should not be used. The important thing to note here is that the product-moment correlation coefficient measures the strength of the linear relationship between two variates. If the relationship is not linear, but is, say, curvilinear, the coefficient can be calculated but there will be distortion and bias in any factors which are calculated from such deceptive coefficients. Note a very important distinction here: the relationships between the variates must be linear, but there must not be a linear dependency between variates: one variate cannot be the sum or the weighted sum of two or more other variates, in effect.

f. Editing: Often one will be tempted to throw away data which do not fall in line with the rest of the material. The guiding principle here is that one can reject it only if one can be sure that he will not be tempted in the future to apply the results to other data which is similarly out of line.

Appendix IV

THE REFERENCE GUIDE TO FACTOR ANALYSIS

INTRODUCTION TO THE REFERENCE GUIDE

A factor analysis provides a description of n variables by a linear combination of m hypothetical factors. The reference guide is designed to help a scientist to obtain such a representation. Each step presents a decision to be made by the user and refers to subsections and appendices of the report which will help him make these decisions.

3. Decide whether to scale the correlation coefficients 3.4

4. Compute the correlation matrix in proper format . . . . . App. 1.A

C. THE FACTOR ANALYSIS

1. Choose the factor analysis technique to be used 4.3

   a. Principal-factor technique . . . . . App. 1.A

   b. Centroid technique . . . . . Refs. 10, 66

2. Decide upon the communality values 4.4

   a. For N > 40, choose unities . . . . . App. 1.A

   b. For N < 40 and for interpretive purposes, choose squared multiple correlations . . . . . 2.7; App. 1.A

   c. For N < 40 and data reduction purposes (preservation of gramian properties 3.3), choose the method of 4.4

D. ROTATION

1. Decide whether to rotate 5.2

   a. If purpose of factor analysis is data reduction: no rotation

   b. If purpose of factor analysis is interpretation: rotation

2. Choose the number of factors to rotate 4.5

3. Select the kind of rotation technique 5.3

   a. Orthogonal rotation (Varimax), if uncorrelated, that is independent factors are hypothesized . . . . . App. 1.B

   b. Oblique rotation (Oblimax), if correlated, that is dependent factors are hypothesized . . . . . App. 1.C

Appendix V

GLOSSARY

Bi-factor solution: a solution, where the variables are described by a general factor, uncorrelated group factors, and a unique factor each.

Biserial correlation coefficient: a bivariate correlation coefficient, where one variable is dichotomized and one variable has quantitative scores.

Centroid solution: a close approximation to the principal-factor solution with considerable saving in labor, where the n variables are described as well by m common and n unique factors.

Common factor: a factor present in more than one variable of a set of variables.

Common-factor space: the space of m common factors.

Communality of a variable: the sum of the squared common factor loadings of the variable; or, the contribution of the common factors to the total unit variance of the variable; or, common-factor variance.

Complete correlation matrix: a correlation matrix with ones in the main diagonal.

Complete factor pattern: a factor pattern which represents the total unit variance of each variable.

Completeness of factorization: the problem of when to stop factoring, that is when to stop extracting factors.

Completeness test: a test to check for completeness of factorization.

Complexity of a variable: the number of common factors involved in the description of a variable.

Contingency coefficient: a bivariate correlation coefficient, where both variables are classified into two or more categories.

Correlation coefficient: the coefficient describing the linear interrelationship of two variables.

Correlation matrix: a real, symmetric square matrix R, whose elements r_{ij} are the correlation c officients between standardized variables Z_i and Z_j.

Covarimin: an oblique rotation method.

Dichotomized variable: a variable which is given by its frequencies in two classes.

Error factor: see specific factor.

Factor: factors are defined as the hypothetical constructs or hypothetical variables in terms of which a variable is linearly represented.

Factor analysis: the analysis of a set of variables into a set of common and unique factors by factoring the correlation matrix of those variables.

Factoring problem: the problem of factoring a given correlation matrix into a factor matrix with an arbitrary reference frame.

Factor loading: same as loading of a factor.

Factor matrix: the matrix of factor loadings.

Factor method: a method to factor a corrrlation matrix in order to obtain a representation of a set of variables in terms of factors.

Factor model: the factor model is given by a set of n equations describing n variables in terms of m common and n unique factors under the assumption that the variables are linearly composed of the factors.

Factor pattern: the set of equations describing a set of n variables in terms of m common and n unique factors; sometimes only the table of factor loadings with the factor designations at the head of the columns are referred to as a pattern.

Factor score: the elements of a factor vector.

Factor solution: a solution to a given factoring problem; often the factor methods are called factor solutions.

Factor structure: a factor structure is a table of correlations between the variables and the factors.

Four-point coefficient: same as ϕ-coefficient.

General factor: a factor present in all variables of a set of variables.

Gramian matrix: a symmetric, positive semidefinite matrix, where a symmetric matrix R is a matrix for which $R = R^T$ holds. R^T represents thereby the matrix with rows and columns of R interchanged, called the transpose of R. Positive semidefiniteness of a matrix is defined as the property of a matrix to have only positive or zero principal minors.

Group factor: a factor present in more than one but not in all variables of a set of variables.

Indeterminacy in factor analysis: referring to the infinitude of factor solutions accounting for the factorization of an observed correlation matrix.

Kaiser-Dickman Method: an oblique rotation method.

Kendall's τ-correlation coefficient: a bivariate correlation coefficient for ranked data.

Loadings of a factor: the coefficients of the factors in the representation of variables by the factors.

Multiple-factor solution: this solution is obtained by transformation (rotation) of a principal-factor or centroid solution according to the principles of simple structure.

Multiple-group solution: a factor solution in which several common factors are extracted in one operation, where these factors can be oblique.

Oblimax: an oblique rotation method.

Oblimin: an oblique rotation method.

Oblique rotation method: the reference frame after rotation is an oblique one.

Observation = measurement = subject = object = individual.

Observed correlation coefficient: a correlation coefficient computed from observed data.

Orthogonal rotation method: the reference frame after rotation is an orthogonal one.

Pattern: same as factor pattern.

Pearson's product-moment correlation coefficient: a bivariate correlation coefficient for quantitative measurements.

ϕ-coefficient: a bivariate correlation coefficient for truely dichotomized variables.

Positive semidefiniteness: see Gramian.

Preferred position of a reference frame: a reference frame for which the factor pattern has a certain prescribed format, where this format can be given in different ways, for example by the simple structure criteria.

Principal component solution: a principal-factor solution of a complete correlation matrix; there are no unique factors.

Principal-factor solution: an orthogonal solution, where the variables are described by m common and n unique factors; the reduced correlation matrix is factored.

Product-moment correlation coefficient: same as Pearson's product-moment correlation coefficient.

Quartimax: an orthogonal rotation method.

Quartimin: an oblique rotation method.

Rank: if N objects are arranged in an order according to some property, which they all possess in a varying degree, the objects are said to be ranked; each object has a rank expressed as a natural number between 1 and N.

Rank of a matrix: the rank of a matrix is the number of rows (or columns) of the largest submatrix whose determinant is not zero.

Reduced correlation matrix: a correlation matrix with communalities in the main diagonal.

Reduced factor pattern: a factor pattern which represents the common factor variance of each variable.

Reference axes: geometrical interpretation of the factors for rotation; the configuration of the reference axes can be oblique or orthogonal.

Reference frame: the frame of reference axes.

Reproduced correlation coefficient: a correlation coefficient reproduced from the pattern of factor loadings.

Residual correlation coefficient: a correlation coefficient computed as the difference between an observed and a corresponding reproduced correlation coefficient.

Residual matrix: a matrix whose entries are the residual correlation coefficients.

Rotation: procedure to re-orient the arbitrary reference axes, determined by the method of factoring the correlation matrix, to some position useful for the interpretation of factors.

Rotational problem: the problem of rotating the arbitrary reference frame, obtained as the result of factoring the correlation matrix, into a preferred position.

Rotation method: same as rotation technique.

Rotation technique: a technique to solve the rotational problem; there are orthogonal and oblique rotation techniques.

Simple structure: a format of the factor pattern, established by Thurstone, as the goal of rotation, observing several criteria.

Spearman's rank correlation: a bivariate correlation coefficient, where both variables are ranked.

Spearman's rank difference method: same as Spearman's rank correlation.

Spearman's ρ-correlation coefficient: same as Spearman's rank correlation.

Specific factor: results from decomposing the uniqueness of a variable into two portions of variance--that due to the particular variable set and that due to error in measurement. Correspondingly two factors are defined: the specific factor and the error factor.

Standardized variable: a variable whose mean is zero and whose standard deviation is one.

Structure: same as factor structure.

Symmetric matrix: see Gramian.

Tetrachoric correlation coefficient: a bivariate correlation coefficient, where both variables are dichotomized.

Thorndike's median ratio coefficient of correlation: a bivariate correlation coefficient for quantitative data.

Total contribution of a factor to the variances of all variables: the sum of squared loadings of all variables on that factor.

Total-factor space: the space of m common and n unique factors.

Trace of a matrix: the sum of diagonal values of a matrix.

Two-factor solution: a solution, where all variables are described by one general factor and one unique factor each.

Uni-factor solution: an orthogonal factor solution, where groups of variables are each described by only one factor.

Unique factor: a factor present in a single variable of a set of variables.

Uniqueness: the contribution of the unique factor of a variable to the unit variance of that variable.

Uniqueness of a solution: the problem referring to discrepancies of two factor solutions due to sampling effects.

Variable: a vector of N observed values where N is the number of observations.

Varimax: an orthogonal rotation method.

Yule's coefficient of association: a bivariate correlation coefficient for dichotomized data.

Yule's coefficient of colligation: a bivariate correlation coefficient for dichotomized data.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Systems Research Laboratories, Inc. Dayton, Ohio	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b GROUP

3. REPORT TITLE

A Survey and Extension of Techniques in Factor Analysis

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report, Feb 64-Feb 65

5. AUTHOR(S) (Last name, first name, initial)

Brand, David H.
Irwin, Larry E.
Ostermann, Elizabeth B.
Donnelly, Thomas G.
Duley, James N.
Shoemacher, Jeff C.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
August 1965	228	66

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
AF30(602)-3300 b. PROJECT NO. 5581 c. Task 558110 d.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) RADC-TR-65-124

10. AVAILABILITY/LIMITATION NOTICES

Qualified requestors may obtain copies of this report from DDC.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Rome Air Development Center Griffiss Air Force Base, New York

13. ABSTRACT

The purpose of the work here was to present the structure of factor analysis to a physical scientist and to extend the structure where it was weakest.

The reference guide in the appendix performs as a dynamic survey of factor analysis by guiding a neophyte factor analyst through an application. Reference is made to expand presentations in the body of the report.

The structure of factor analysis has been extended in the following areas: effects on the number of observations, sampling effect, interpretation of factors and communality.

DD FORM JAN 64 1473

UNCLASSIFIED

Security Classification

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Factor Analysis Mathematical Analysis Statistical Processes						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

UNCLASSIFIED

Security Classification